Springer Undergraduate Mathematics Series

S U M S

Asuman Güven Aksoy

# Fundamentals of Real and Complex Analysis

Springer

# Springer Undergraduate Mathematics Series

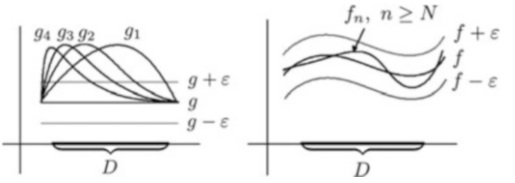
**Editor-in-Chief**
Endre Süli, Oxford, UK

**Series Editors**
Mark A. J. Chaplain, St. Andrews, UK
Angus Macintyre, Edinburgh, UK
Shahn Majid, London, UK
Nicole Snashall, Leicester, UK
Michael R. Tehranchi, Cambridge, UK

The Springer Undergraduate Mathematics Series (SUMS) is a series designed for undergraduates in mathematics and the sciences worldwide. From core foundational material to final year topics, SUMS books take a fresh and modern approach. Textual explanations are supported by a wealth of examples, problems and fully-worked solutions, with particular attention paid to universal areas of difficulty. These practical and concise texts are designed for a one- or two-semester course but the self-study approach makes them ideal for independent use.

Asuman Güven Aksoy

# Fundamentals of Real and Complex Analysis




Springer

Asuman Güven Aksoy
Department of Mathematical Sciences
Claremont McKenna College
Claremont, CA, USA

ISSN 1615-2085 ISSN 2197-4144 (electronic)
Springer Undergraduate Mathematics Series
ISBN 978-3-031-54830-7 ISBN 978-3-031-54831-4 (eBook)
https://doi.org/10.1007/978-3-031-54831-4

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable.

To my grandchildren

Nora, Aydın

# Preface

Real and complex analysis are fundamental topics in understanding mathematics at both undergraduate and graduate levels. While real and complex analysis are traditionally separated, this book unites these subjects after developing the basic techniques to understand them. The few texts which do unify real and complex analysis, such as Rudin's excellent "Real and Complex Analysis," often presume readers begin with a high level of mathematical sophistication. In contrast, this book aims to accommodate a larger population of readers, especially advanced undergraduates, and emphasizes the strong connection between various branches of analysis.

We aim to present different subareas of analysis as interconnected, rather than separate disciplines in order to give the reader understanding of analysis as a whole. Furthermore, this book is geared toward those interested in continuing further in mathematics, whether you are an advancing math major or a quantitative scientist, or simply an interested person wanting to become well-versed in advanced mathematics beyond the level of popularizations. We include ample examples, exercises, and copious references in order to guide those who wish to go deeper into particular topics.

Many of the properties of the field of real numbers $\mathbb{R}$ hold in the field of complex numbers $\mathbb{C}$ as well. There are some truly remarkable differences. For example, many results in real analysis make use of the fact that $\mathbb{R}$ is totally ordered, $\mathbb{C}$ loses order but gains a lot of properties such as well-behaved derivatives. A real valued function could be differentiated once, but there is no guarantee it can be differentiated twice; in contrast, if we have a complex valued differentiable function, then it can be differentiated many times. Although exponential function $e^x$ is one-to-one, $e^z$ is not, in fact it is a periodic function, $\log_e x$ is a single valued function, but $\ln z$ is multi-valued. Writing $\sin x = 5$ makes no sense, but if we replace $x$ by $z$ and write $\sin z = 5$ makes perfect sense.

The book is organized as follows. In Chapter 1 we begin our exploration of advanced mathematics with foundational topics. Starting with set theory, we look at the differences between finite and infinite sets and lay the groundwork for analysis and topology. We introduce equivalence relations and quotient sets and congruences, prompting a discussion of the field axioms. We meet fields, both finite and infinite, before turning our attention to the properties of the field of real numbers, including their completeness and topology. We also present in this chapter a review of some important facts about infinite

sequences and series, building on the reader's experience of calculus to develop the beginning of analysis. With the topological concepts developed earlier, we study continuity, differentiability, and integrability of real valued functions. Properties of continuous functions on compact and connected sets, the mean value theorem, the fundamental theorem of calculus, and the inverse function theorem are all covered. We also touch upon the concept of sets of measure zero and Lebesgue's idea of integrability, which will be studied in detail in the next chapter. Thus, Chapter 1 can be seen as a "toolbox" for higher mathematics. Its contents are easily accessible but not elementary.

In Chapter 2, we focus on topics in real analysis. We start with the concepts of metric, inner product, and normed spaces. Since research in $p$-adic analysis is very prominent not only in analysis but also in number theory and mathematical physics, we examine the $p$-adic completion of the rational numbers relative to a prime $p$. We also study Baire's theorem which is central to functional analysis. We prove Baire's theorem for the reals, completing the discussion from Chapter 1 about the size of the real numbers. The second section in this chapter is on the Banach fixed point theorem, whose utility in solving systems of linear equations as well as its applications to differential and integral equations is illustrated with numerous examples. The topics chosen here are not only of theoretical value, they provide essential knowledge for anyone interested in applied mathematics. We then cover modes of convergence and approximation by polynomials. We emphasize the differences between uniform and pointwise convergence and their effect on the continuity, differentiability, and integrability of sequences or series of functions. Next, we ask the question whether there are any "useful" dense subspaces of the space of continuous functions on a closed and bounded interval, leading us to the Weierstrass approximation theorem, which we prove using Bernstein polynomials. Another proof of the Weierstrass approximation theorem for trigonometric polynomials is mentioned in a subsequent section on Fourier series.

One of the unique features of Chapter 2 is its treatment of functional equations, which is often neglected in most analysis texts. In the section on Fourier series, we start with historical background, then quickly move on to orthogonality relations and emphasize convergence issues. Finally, our discussion shifts to the continuous analog, then to the Fourier Transform, and its inversion, including convolution and the relationship between the Fourier transform and the Laplacian. We introduce Lebesgue measurable sets and functions while stating Littlewood's three principles. Here we will also encounter fantastic examples that stretch the imagination, like the Cantor set and the Cantor-Lebesgue function. The Lebesgue measure will allow us to define the correct notion of integration, through which we will be able to integrate far more functions as well as prove that the integral of the limit of functions is equal to limit of the integrals (the Lebesgue-dominated convergence theorem). We conclude this section with the Banach-Tarski paradox, which roughly shows us a way to take a ball, decompose it into a finite number of pieces, then reassemble it into two balls identical to the original.

In Chapter 3, we cover the most essential topics in complex analysis. Starting with a geometric introduction to the complex plane, we study holomorphic functions, complex power series, conformal mappings, and the Riemann mapping theorem. We focus on the power and significance of Cauchy's theorem, the centerpiece of complex analysis. We emphasize applications of this theorem through Cauchy's integral formula and residue theorem. Even though these concepts date back to the nineteenth century, we illustrate their power when we briefly glimpse into the Bieberbach conjecture. In the decades that the Bieberbach conjecture stood unsolved, mathematicians discovered many properties of univalent functions. In fact, most of the theory of univalent functions and geometric function theory arose from these efforts. The significance of the Bieberbach conjecture lies not only with its solution, but also the theory developed to solve it.

My hope is this book will be useful to anyone interested in upper division mathematics. Textbook uses aside, this book is meant as a coherent response to a motivated student who, wanting to truly understand analysis, asks "Can you please get to the point?"

Asuman Güven Aksoy
Claremont, CA, USA

# Acknowledgments

The results in this book belong to the common heritage of mathematics. Needless to say, I am indebted intellectually to the many mathematicians who contributed to the creation of real and complex analysis. Even though certain theorems are designated with the usual proper names, I have otherwise made no special effort to attribute theorems and proofs.

My special thanks goes to my colleague and dear friend Sam Nelson from Claremont McKenna College. He has been involved with this project since its inception and has created nearly all illustrations in this book. Many many thanks! I also like to thank two graduate students: Chris Donnay of Ohio State University and Daniel Akech of Claremont Graduate University. They both read the entire text and made many valuable corrections. Finally, I thank Elizabeth Loew for her wonderful work as an editor. I am grateful to my best friend Ercüment G. Aksoy for the constant support.

Asuman G. Aksoy
Claremont, California
December 2023

# Contents

**Preface** **vii**

**1 Introductory Analysis** **1**
1.1 Set Theory . . . 1
1.2 Number Systems . . . 22
1.3 Completeness and the Real Number System . . . 28
1.4 Sequences and Series . . . 38
1.5 Topology of the Real Line . . . 63
1.6 Continuous Functions . . . 84
1.7 Differentiability on $\mathbb{R}$ . . . 98
1.8 The Riemann Integral . . . 111

**2 Real Analysis** **127**
2.1 Metric, Normed, and Inner Product Spaces . . . 127
2.2 Fixed Point Theorems and Applications . . . 173
2.3 Modes of Convergence . . . 191
2.4 Approximation by Polynomials . . . 203
2.5 Functional Equations . . . 213
2.6 Fourier Series . . . 223
2.7 Lebesgue Measure and Integration . . . 240
2.8 Banach–Tarski Paradox . . . 272

**3 Complex Analysis** **277**
3.1 The Complex Plane . . . 277
3.2 Holomorphic Functions . . . 289
3.3 Power Series . . . 296
3.4 Some Holomorphic Functions . . . 304
3.5 Conformal Mappings . . . 315
3.6 Integration in the Complex Plane . . . 331
3.7 Cauchy's Theorem . . . 337
3.8 Cauchy's Formulae . . . 350
3.9 Laurent Expansion and Singularities . . . 359
3.10 The Bieberbach Conjecture . . . 375

**About the Author** **383**

**Bibliography** **385**

**Index** **391**

# Chapter 1

# Introductory Analysis

> If now to any $x$ there corresponds a unique, finite $y$... Then $y$ is called a function of $x$ for this interval... This definition does not require a common rule for different parts of the curve; one can imagine the curve as being composed of the most heterogeneous components or as being drawn without following any law.
>
> *Dirichlet 1837*

## 1.1 Set Theory

Some ancient mathematicians did not consider zero to be a number. It was not that they had not thought of zero; rather, they considered numbers to be lengths of line segments, i.e., geometric extensions in space, and a line segment of length zero seems like hardly a line segment at all. This lack of zero meant that ancient mathematics lacked the co ncept of an *empty set*, i.e., a set without any elements. This in turn had consequences for ancient logic, leading to certain invalid arguments being accepted as valid. Modern logic arises from *set theory*, and indeed all of mathematics has set theory as its fundamental base. For further reading on the topics of this chapter, we recommend the following references: [3, 9, 17, 28, 29, 42, 43, 46, 52].

### Sets

A *set* is an unordered list or collection $X$ of things called *elements*. We can define a set by explicitly listing its elements, like

$$X = \{1, 2, 3, \pi\},$$

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

A. G. Aksoy, *Fundamentals of Real and Complex Analysis*, Springer Undergraduate Mathematics Series, https://doi.org/10.1007/978-3-031-54831-4_1

or by giving a defining property or properties, like

$$X = \{x \mid x \text{ is an even integer}\}.$$

The order of the elements listed in a set does not matter;

$$\{1, 2, 3\} = \{1, 3, 2\} = \{2, 3, 1\}.$$

Multiple copies of elements are ignored in sets, so the set $\{1, 2, 3, 1\}$ is just the set $\{1, 2, 3\}$. We can also define *multisets* or *sets with multiplicity* in which each element has a multiplicity or the number of times it appears, e.g., $\{1, 1, 1, 2, 2, 3\}$. We generally use capital letters like $X$ and $Y$ for sets and lowercase letters like $x$ and $y$ for elements of sets. The symbol $\in$ means "is an element of," so $x \in X$ means that $x$ is an element of the set $X$.

To be a set, $X$ must have the property that the question of whether a given thing belongs to $X$ is always decidable. This membership-decidability condition is necessary to avoid *Russell's paradox*, named after the great mathematician and philosopher Bertrand Russell.

To see why, temporarily suppose that every collection of elements is a set (this approach is called *naive set theory*). Say a set is *self-containing* if it contains a copy of itself as an element, i.e., $X$ is self-containing if $X \in X$. For example, the set of all sets is self-containing, while the set $\{1, 2, 3\}$ is not self-containing since its elements are numbers, not sets. Russell then asked "What about the set of all non-self-containing sets, $Y = \{X \mid X \notin X\}$?" In particular, is $Y$ self-containing or not? Well, if $Y \in Y$, then $Y$ is self-containing, but $Y$ is defined to be the set of only those sets that are not self-containing, so it seems clear that $Y \notin Y$. But this then says that $Y$ *is* in $Y$, since $Y$ is the set of all non-self-containing sets, so $Y \in Y$. In fact, we have $Y \in Y$ if and only if $Y \notin Y$, a contradiction.

This shows that it is necessary to make a distinction between *sets*, where issues of inclusion are always decidable, and *proper classes*, where decidability of inclusion may fail for some cases. The collection of all non-self-containing collections is a proper class, not a set, and there is no set of all non-self-containing sets.

A set $S$ is a *subset* of a set $X$, denoted $S \subset X$, if for every $x \in S$, we have $x \in X$; that is, $S$ is a subset of $X$ if $x \in S \Rightarrow x \in X$. Note that a set is always a subset of itself. A subset $S$ of $X$ which is not equal to $X$ is called a *proper subset*, written $S \subsetneq X$. If $S \subsetneq X$, then there is at least one element $x \in X$ such that $x \notin S$. For instance,

$$\{1, 2, 3\} \subsetneq \{0, 1, 2, 3, 4, 5\}.$$

Specifically, the statement "$X \subset Y$" ("$X$ is a subset of $Y$") means "for any $x$, if $x \in X$, then $x \in Y$."

There is a unique set with no elements, called the *empty set*, denoted $\emptyset = \{\}$. For every set $X$, we have $\emptyset \subset X$, since $\emptyset \not\subset X$ would mean there is an element $x \in \emptyset$ such that $x \notin X$, and this condition is never satisfied.

The *power set* of a set $A$ is the set of all subsets of $A$, i.e.,

$$\mathcal{P}(A) = \{S \mid S \subset A\}.$$

The power set of the set $A = \{1, 2, 3\}$ is the set

$$\mathcal{P}(A) = \{\{1, 2, 3\}\, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1\}, \{2\}, \{3\}, \emptyset\}.$$

One useful way to avoid the issues that can arise when dealing with proper classes is to select an appropriately large set to be our "universe" and restrict ourselves to subsets of this universal set, i.e., elements of the power set of our universal set. For instance, we might only be interested in integer solutions for a given problem, in which case we could declare the set of integers to be our universal set.

## Set Operations

Given two sets $A$ and $B$, we can combine $A$ and $B$ into new sets in several ways:

- **Union.** The *union* of $A$ and $B$ denoted $A \cup B$ is the set in Figure 1.1.

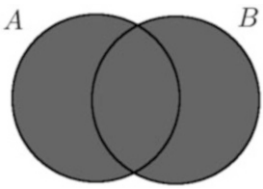



Figure 1.1: $A \cup B = \{x \mid x \in A \text{ or } x \in B\}$

For example,

$$\{1, 2, 3, 4\} \cup \{3, 4, 5\} = \{1, 2, 3, 4, 5\}.$$

- **Intersection.** The *intersection* of $A$ and $B$ denoted $A \cap B$ is the set in Figure 1.2.

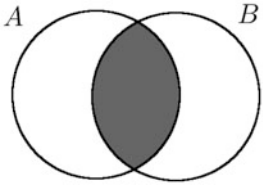



Figure 1.2: $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$

For example,

$$\{1, 2, 3, 4\} \cap \{3, 4, 5\} = \{3, 4\}.$$

Two sets $A$ and $B$ are called *disjoint* if their intersection is empty, i.e., if $A \cap B = \emptyset$. For example, $\{1, 2, 3\}$ and $\{4, 5, 6\}$ are disjoint sets.

- **Setwise Difference.** The *setwise difference* of $A$ and $B$ denoted $A \setminus B$ and read "$A$ minus $B$" is shown in Figure 1.3.

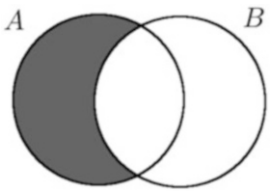



Figure 1.3: $A \setminus B = \{x \mid x \in A \text{ and } x \notin B\}$

For example,

$$\{1, 2, 3, 4, 5\} \setminus \{2, 5, 6, 7\} = \{1, 3, 4\}.$$

- **Complement.** If $B \subset A$, then the setwise difference $A \setminus B$ is called the *complement* of $B$ in $A$. For example, the complement of $\{1, 2, 3\}$ in $\{1, 2, 3, 4, 5\}$ is $\{4, 5\}$. In particular, if we are working within a universal set $X$, then the complement of a set $S \subset X$ is the set of all things outside of $S$.

- **Cartesian Product.** The *Cartesian product* of the sets $A$ and $B$ denoted $A \times B$ is the set of all ordered pairs whose first element is from $A$ and whose second element is from $B$, i.e.,

$$A \times B = \{(a, b) \mid a \in A \text{ and } b \in B\}.$$

For example, if $A = \{1, 2, 3\}$ and $B = \{a, b\}$, then

$$A \times B = \left\{ \begin{array}{lll} (1, a), & (2, a), & (3, a) \\ (1, b), & (2, b), & (3, b) \end{array} \right\}.$$

The Cartesian product of a set $X$ with itself, $X \times X$, is often abbreviated $X^2$. We can also define repeated Cartesian products – for example, the set $X \times X \times X = X^3$ is the set of ordered triples of elements of $X$, and $X^n$ is the set of ordered $n$-tuples (i.e., ordered lists of length $n$) of elements of $X$. For example, the set $\mathbb{R}$ of real numbers forms a line, while the set $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$ forms a plane, etc.

Set theory is at the core of everything we do in mathematics. For instance, multiplication arises from the Cartesian product – in the example above, the product of a two-element set with a three-element set is a $2 \times 3 = 6$ element set. In the next section we will see how logic itself arises from set theory. For the remainder of this section, we will see one of the most important ideas in all of mathematics, the idea of a *function*.

## Functions

Let $X$ and $Y$ be sets. A *function* or *mapping* $f : X \to Y$ from $X$ to $Y$ is a rule $f$ which takes input values $x$ from $X$ and assigns to each $x$ a *unique* output value $f(x) \in Y$, as seen in Figure 1.4.

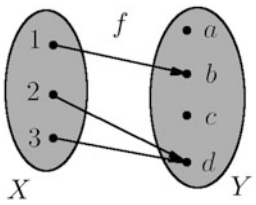



Figure 1.4: A function $f$

A function can be conceptualized as machine turning elements of $X$ into elements of $Y$. Formally speaking, a function $f$ is a subset of the Cartesian product $X \times Y$ where each element $x \in X$ appears exactly once in the first position so that each $x$ gets an unambiguously defined $f(x)$. For example, let $X = \{1, 2, 3\}$ and $Y = \{a, b, c, d\}$. Then the Cartesian product is

$$X \times Y = \{(1,a),(1,b),(1,c),(1,d),(2,a),(2,b),(2,c),(2,d),(3,a),(3,b),(3,c),(3,d)\}.$$

On the other hand, there is a function $f : X \to Y$ defined by $f(1) = b$, $f(2) = d$, and $f(3) = d$; we can identify this set with the subset

$$f = \{(1,b),(2,d),(3,d)\} \subset X \times Y. \tag{1.1}$$

Almost everything in mathematics involves functions. Calculus is largely the study of functions $f : \mathbb{R} \to \mathbb{R}$, where $\mathbb{R}$ is the set of *real numbers*; multivariable calculus involves the study of functions of the form $f : \mathbb{R}^n \to \mathbb{R}^m$. Linear algebra, abstract algebra, topology, geometry, and analysis all come down more or less to the study of functions satisfying certain properties.

One way to define a function is with a formula, e.g., $f : \mathbb{R} \to \mathbb{R}$ is given by

$$f(x) = x^3 - \sin x \quad \text{or} \quad f(x) = \begin{cases} x^2 - 1 & : x \leq 0 \\ \sqrt{x+1} & : x > 0. \end{cases}$$

A function $f : \mathbb{R} \to \mathbb{R}$ can be defined with a graph, i.e., a curve that does not intersect any vertical line more than once. Then each point has coordinates $(x, f(x))$. From the graph we can see very clearly that the function is a subset of the cartesian product space $\mathbb{R} \times \mathbb{R}$, i.e., the $xy$-plane.

Let $f : X \to Y$ be a function. The set of input values $X$ is called the *domain* of the function $f$, and the set $Y$ of possible output values is called the *codomain* of $f$. The *image* of $f$ is the set

$$\mathrm{Im}(f) = f(X) = \{y \in Y \mid y = f(x) \text{ for some } x \in X\}.$$

The image of a function is always a subset, not necessarily proper, of the codomain. Note that some authors use the term "range" to refer to the image of $f$, while others use the term "range" to mean the codomain of $f$. We will prefer to avoid using the term "range" altogether to avoid confusion.

For any subset $S \subset Y$, the *preimage* of $S$ is

$$f^{-1}(S) = \{x \in X \mid f(x) \in S\},$$

the set of all elements of $X$ that get sent into $S$ by $f$. In our example of $f$ in (1.1) above, we have $f(X) = \{b, d\} \subset Y$ but $f(X) \neq Y$; $f^{-1}(\{b\}) = \{1\}$, $f^{-1}(\{d\}) = \{2, 3\}$, and $f^{-1}(\{a\}) = \emptyset$. Despite the use of $f^{-1}$ notation, this has nothing to do whether or not $f$ is invertible.

A function $f : X \to Y$ is *injective* or *one-to-one* if no two input values give us the same output value. More formally, $f$ is injective if

$$f(x_1) = f(x_2) \Rightarrow x_1 = x_2.$$

See Figure 1.5.

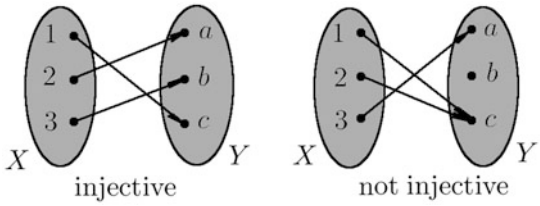



Figure 1.5: Injectivity

**Example 1.** We can prove that a given function is injective using the criterion $f(x_1) = f(x_2) \Rightarrow x_1 = x_2$ by supposing that $f(x_1) = f(x_2)$ and showing that $x_1 = x_2$. For example, we claim that the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^3 - 1$ is injective. To prove it, suppose $f(x_1) = f(x_2)$, i.e., $x_1^3 - 1 = x_2^3 - 1$; then we have

$$x_1^3 - 1 = x_2^3 - 1 \iff x_1^3 = x_2^3 \iff x_1 = x_2,$$

and we have $f(x_1) = f(x_2) \Rightarrow x_1 = x_2$.

**Observation 1.** Another equivalent definition of injectivity is that $f$ is injective if and only if the preimage $f^{-1}(\{y\})$ of every single element set has at most one element.

A function $f : X \to Y$ is *surjective* or *onto* if every potential output value in $Y$ gets hit by $f$. More formally, $f$ is surjective if

$$y \in Y \Rightarrow y = f(x) \text{ for some } x \in X.$$

In particular, $f$ is surjective if $f(X) = Y$, i.e., if the image and codomain are the same. See Figure 1.6.

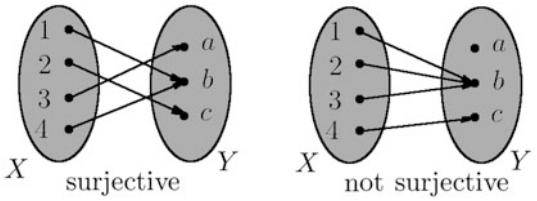



Figure 1.6: Surjectivity

**Observation 2.** A function $f : X \to Y$ is surjective if and only if no $f^{-1}(\{y\})$ is empty, or equivalently $f^{-1}(S) = \emptyset$ implies $S = \emptyset$.

A function that is both injective and surjective is called *bijective*; a bijective function is called a *bijection*. A bijection is essentially a relabeling of elements of $Y$ with labels in $X$. Note that functions in general can be injective, surjective, bijective, or neither.

**Example 2.** For any set $X$, the *identity function* on $X$, $\mathrm{Id}_X : X \to X$, is given by $\mathrm{Id}_X(x) = x$ for all $x \in X$. That is, $\mathrm{Id}_X$ is the function that just hands you back whatever input value you give it as output. The identity function is a bijection.

If $f : X \to Y$ is a bijection, then for every $y \in Y$ the preimage set $f^{-1}(\{y\})$ is nonempty and contains a unique element $x$; in this situation, we can define the *inverse function* $f^{-1} : Y \to X$ by setting $f^{-1}(y) = x$, where $f^{-1}(\{y\}) = \{x\}$.

If $f : X \to Y$ and $g : Y \to Z$, there is a function $g \circ f : X \to Z$ called the *composite* of $f$ and $g$ defined by $(g \circ f)(x) = g(f(x))$. See Figure 1.7.

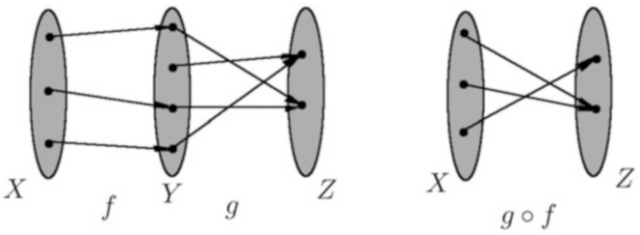



Figure 1.7: Composition of $g$ by $f$

**Example 3.** For functions given by formulas, composing $f(x)$ with $g(x)$ means evaluating $g$ at $f(x)$, i.e., "plugging in" $f(x)$ into $g(x)$, so that

$$(g \circ f)(x) = g(f(x)).$$

For instance, if $g(x) = x^2 + 3x - 1$ and $f(x) = x + 1$, then

$$\begin{aligned} g(f(x)) &= f(x)^2 + 3f(x) - 1 = (x+1)^2 + 3(x+1) - 1 \\ &= x^2 + 2x + 1 + 3x + 3 - 1 = x^2 + 5x + 3. \end{aligned}$$

**Example 4.** Suppose $f : X \to Y$ and $g : Y \to X$ satisfy $(g \circ f)(x) = x$ and $(f \circ g)(y) = y$ for all $x \in X$ and $y \in Y$. Then we say $g$ is the *inverse function* of $f$ and write $g = f^{-1}$. Note that $f : X \to Y$ has an inverse function if and only if $f$ is bijective, i.e., if and only if the preimage set $f^{-1}(\{y\})$ of each element of $Y$ has exactly one element $x$; then the inverse function simply sends each element $y$ to its preimage value $x$.

If $f : X \to Y$, $g : Y \to Z$, and $h = g \circ f : X \to Z$, we say that $h$ *factors through* $g$ (or $f$ or $Y$). This is often expressed in the form of a diagram, e.g., Figure 1.8.

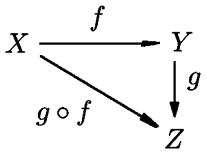



Figure 1.8: $g \circ f$ factors through $f$ or $g$

In particular, a function $f : X \to Y$ has an inverse if and only if the identity map $I_X : X \to X$ factors through $f$ (Figure 1.9).

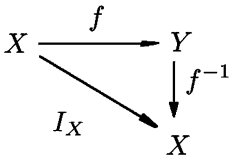



Figure 1.9: Factorization of the identity map

## Induction

One important "tool" of the trade which one encounters quite often is the use of induction arguments. Induction is used in conjunction with the natural numbers $\mathbb{N}$ or sometimes with $\{0\} \cup \mathbb{N}$. The general principle behind induction is as follows:
Let $P(n)$ be a proposition about $n$. Then $P(n)$ is true for all $n \in \mathbb{N}$, provided that:

a) $P(1)$ is true.

b) For each $k \in \mathbb{N}$, if $P(k)$ is true, then $P(k+1)$ is true.

We refer the verification of part a) above as the *basis for induction* and part b) as the *inductive step*. The assumption that $P(k)$ is true is known as the *induction hypothesis*. The following is a simple example of a proof by induction.

**Example 5.** Prove that

$$1 + 2 + 3 + \cdots + n = \frac{1}{2}n(n+1)$$

for every natural number $n$.

*Proof.* Let $P(n)$ be the statement $1 + 2 + 3 + \cdots + n = \frac{1}{2}n(n+1)$. Then $P(1) = 1 = \frac{1}{2}(1)(1+1)$ is true, and this establishes the basis for induction. Next we suppose that $P(k)$ is true, where $k \in \mathbb{N}$ (induction hypothesis). That is, we assume

$$1 + 2 + 3 + \cdots + k = \frac{1}{2}k(k+1).$$

Since we wish to show $P(k+1)$ is true, we add $k+1$ to both sides to obtain

$$\begin{aligned} 1 + 2 + 3 + \cdots + k + (k+1) &= \frac{1}{2}k(k+1) + (k+1) \\ &= \frac{1}{2}\left[k(k+1) + 2(k+1)\right] \\ &= \frac{1}{2}(k+1)(k+2) \\ &= \frac{1}{2}(k+1)\left[(k+1)+1\right]. \end{aligned}$$

Thus $P(k+1)$ is true whenever $P(k)$ is true, and by induction we conclude that $P(n)$ is true for all $n$. □

**Remark 1.** There is a generalization of the principle of mathematical induction that enables us to conclude that a given statement is true for all natural numbers sufficiently large. In this case, let $P(n)$ be a proposition about $n$ and $a \in \mathbb{N}$. Suppose that:

a) $P(a)$ is true.

b) For each $k \geq a$, if $P(k)$ is true, then $P(k+1)$ is true.

Then $P(n)$ is true for all $n \geq a$.

**Example 6.** Suppose we want to show for all $n \geq 5$, $4n < 2^n$. We again use induction, but base case is now $k = 5$ since $4k = 20 < 32 = 2^k$. By the inductive hypothesis we assume for some $k \geq 5$ we have $4k < 2^k$. Next we consider $P(k+1)$ and write

$$4(k+1) = 4k + 4 < 2^k + 4 = 2^{k+1}.$$

Thus by induction we have that $4n < 2^n$ for all $n \geq 5$.

## Cardinality

We start by defining *finite sets*. A set $X$ is finite if $X$ is empty or there exists $n \in \mathbb{N}$ such that there is a bijection $f : X \to \{1, 2, \ldots, n\}$, where $\{1, 2, \ldots, n\}$ is the set of all natural numbers less than or equal to $n$. If $X$ is a finite set, then the *cardinality* of $X$, denoted $|X|$ or $card(X)$, is the number of elements in $X$. For example,

$$|\{a, b, c, d\}| = 4.$$

Note that the cardinality of a finite set is a natural number. Indeed, the set of natural numbers is precisely the set of cardinalities of finite sets, provided we include $0 = |\emptyset|$. We can start with some useful observations:

**Observation 3.** There is an injective map $f : X \to Y$ if and only if $|X| \leq |Y|$. Why? Well, define a map $f : X \to Y$ by choosing an element of $Y$ for each element of $X$, and since we are trying to make $f$ injective, we want to avoid reusing elements. If $|Y| < |X|$, then we will run out of elements of $Y$ before we have a complete function, so we will have to reuse some elements. On the other hand, if $|Y| \geq |X|$, then we will be able to avoid reusing elements of $Y$ and possibly still have elements of $|Y|$ to spare. Thus, there exists an injective map $f : X \to Y$ if and only if $|X| \leq |Y|$.

**Observation 4.** There is a surjective map $f : X \to Y$ if and only if $|X| \geq |Y|$. Again imagine going through the elements of $X$, assigning an element of $Y$ to each. We will have a complete function when all elements of $X$ have an $f(x)$ assigned, and if $|X| < |Y|$, then we will run out elements of $X$ before all of the elements of $Y$ are used. On the other hand, if $|X| \geq |Y|$, then there are enough elements $x \in X$ to assign every $y \in Y$ to some $x$. Thus, there exists a surjective function $f : X \to Y$ if and only if $|X| \geq |Y|$.

Putting these two together, we have the following definition:

**Definition 1.** Two sets $X$ and $Y$ have the same cardinality if and only if there exists a bijection $f : X \to Y$.

This definition provides a nice example of how intuition can go wrong. For example, if asked to provide a definition in terms of set theory for when one set is larger than another, many of us might suggest that $|X| < |Y|$ if $X \subsetneq Y$. Certainly it works this way for finite sets; however, infinite sets have the counter-intuitive property that they can be put in bijective correspondence with proper subsets of themselves. That is, for infinite sets, a proper subset can be the same size as the whole set. Indeed, this is the very definition of what it means for a set to be infinite. In particular, we have the following definition:

**Definition 2.** A set $X$ is *infinite* if there is a proper subset $S \subsetneq X$ and a bijection $f : X \to S$.

**Example 7.** The set of natural numbers

$$\mathbb{N} = \{1, 2, 3, 4, \ldots, n, \ldots\}$$

is infinite since there is a bijection $f : \mathbb{N} \to 2\mathbb{N}$ between the set of all natural numbers $\mathbb{N}$ and the set of even natural numbers $2\mathbb{N}$ given by $f(n) = 2n$, where $n \in \mathbb{N}$. The function $f$ is injective since $2n = 2n'$ implies $n = n'$ and $f$ is surjective since every even natural number is $2n$ for some natural number $n$. More visually, we can see the correspondence with a table:

| $\mathbb{N}$ | 1 | 2 | 3 | 4 | 5 | 6 | ... |
|---|---|---|---|---|---|---|---|
| $2\mathbb{N}$ | 2 | 4 | 6 | 8 | 10 | 12 | ... |

.

Thus, there are exactly as many even natural numbers as natural numbers, and the set of natural numbers is infinite. Similarly, there are exactly as many natural numbers as there are integers, which we can see most easily by listing a correspondence:

| $\mathbb{N}$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | ... |
|---|---|---|---|---|---|---|---|---|---|---|
| $\mathbb{Z}$ | 0 | $-1$ | 1 | $-2$ | 2 | $-3$ | 3 | $-4$ | 4 | ... |

or, for those preferring a formula, define

$$f(n) = \begin{cases} -\dfrac{n}{2} & n \text{ even} \\ \dfrac{n-1}{2} & n \text{ odd.} \end{cases}$$

While this property is very counter-intuitive, we should not expect to have much natural intuition about infinity since it is something we have never (and can never) directly experience. Nonetheless, with careful mathematical reasoning, we can learn many interesting facts about infinity.

The following theorem is very useful because it allows us to conclude two sets have the same cardinality without explicitly constructing a bijection between the sets. This theorem is sometimes called the Schröder–Bernstein theorem. E. Schröder and F. Bernstein gave independent proofs of this theorem.

**Theorem 1** (Schröder–Bernstein Theorem)**.** *Let $X$ and $Y$ be nonempty sets. If there exist a one-to-one map $f : X \to Y$, from $X$ into $Y$, and a one-to-one map $g : Y \to X$, from $Y$ into $X$, then there is a map $h : X \to Y$ that is both one-to-one and onto.*

*Proof.* Recall that the image of $f$ is defined as

$$f(X) = \{y \in Y : \ y = f(x) \ \text{ for some } x \in X\}.$$

Because $f$ is not necessarily onto, the image $f(X)$ may not be all of $Y$. Let $x \in X$ be arbitrary, and let $C_x$ be the list of all elements of the form

$$C_x = \{x,\, g^{-1}(x),\, f^{-1} \circ g^{-1}(x),\, g^{-1} \circ f^{-1} \circ g^{-1}(x), \dots\}.$$

The elements of this sequence are called *predecessors* of $x$. Notice that since we started with $x \in X$, then if $g^{-1}(x)$ exists, in other words if $x$ is in the image of $g$, then $g^{-1}(x) \in Y$. For each $x \in X$, one of the three following possibilities happens:

a) The list $C_x$ is infinite. In this case $x$ has infinitely many predecessors and the corresponding subset of $X$ will be denoted by $X_1$.

b) The last term in the list is an element of $X$. The corresponding subset of $X$ will be denoted by $X_2$.

c) The last term in the list is an element of $Y$. Let $X_3$ denote the corresponding subset of $X$.

If the last term in the list is an element of $X$, then the last term is of the form $y = f^{-1} \circ g^{-1} \circ \cdots \circ g^{-1}(x)$ or $y = x$, and $g^{-1}(y)$ does not exist. In this case $C_x$ stops in $X$. This explains case 2 above. If the last term in the list is an element of $Y$, that is, the last term is of the form $w = g^{-1}(x)$ or $w = g^{-1} \circ f^{-1} \circ \cdots \circ g^{-1}(x)$ and $w$ is not in the image of $f$, then we are describing case 3 above. Just like the subsets $X_1, X_2$, and $X_3$ described above, we define the corresponding subsets $Y_1, Y_2$, and $Y_3$ as follows:

a) $Y_1 = \{y \in Y : \ y \text{ has infinitely many predecessors}\}$.

b) $Y_2 = \{y \in Y : \ \text{the predecessors of } y \ \text{ stop in } X\}$.

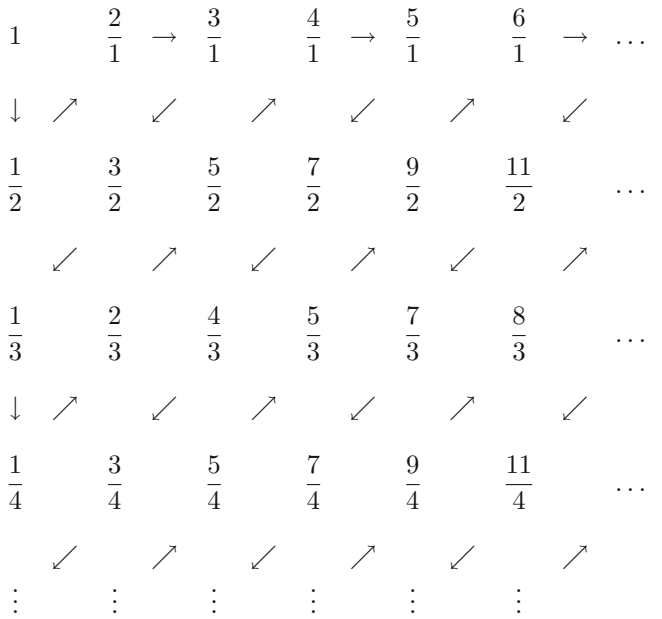
c) $Y_3 = \{y \in Y : \ \text{the predecessors of } y \ \text{ stop in } Y\}$.

Now observe that $X_1, X_2, X_3$ and $Y_1, Y_2, Y_3$ partition $X$ and $Y$, respectively, and the mappings $f : X_1 \to Y_1$, $g : Y_1 \to X_1$ are both bijections. The same is true for $f : X_2 \to Y_2$, $g : Y_2 \to X_2$ and $f : X_3 \to Y_3$, $g : Y_3 \to X_3$. □

**Definition 3.** A set $X$ has cardinality $\aleph_0$ (pronounced as "aleph null") if there is a bijection between the sets $X$ and $\mathbb{N}$. Naturally the set $\mathbb{N}$ has cardinality $\aleph_0$.

**Definition 4.** A set is called *countably infinite* if it has cardinality $\aleph_0$, that is, if it is in one-to-one correspondence with natural numbers. The term "countable" or "denumerable" is also used by some to refer to infinite sets that are in one-to-one correspondence with $\mathbb{N}$, while others include finite sets when they say countable. We will say *countable* if it is finite or countably infinite.

Let $\mathbb{Q}_+$ denote the positive rational numbers. Then we claim that $\mathbb{Q}_+$ has cardinality $\aleph_0$. Even though in the following we do not write down explicitly what the bijection between $\mathbb{Q}_+$ and $\mathbb{N}$ looks like, we understand the idea of bijection by the following *diagonalization* argument. We write all the positive fractions in a grid with all fractions with denominator 1 in the first row, all fractions with denominator 2 in the second row, all the fractions with denominator 3 in the third row, etc. Next go through row by row, and remove all fractions that are not written in the lowest terms. Then start with the upper left-hand corner, and draw a path through all the remaining numbers as shown in the following diagram below:



We can count along the path we drew, assigning $1/1 \to 1$, $1/2 \to 2, 2/1 \to 3, 3/1 \to 4$, etc. This is the bijection we need.

**Remark 2.** It is easy to find an injection from $\mathbb{N}$ to $\mathbb{Q}_+$; simply send $n$ to $n$. If one can find an injection from $\mathbb{Q}_+$ to $\mathbb{N}$ (try this), the Schröder–Bernstein theorem gives a bijection. Then we can say $\mathbb{Q}_+$ has cardinality $\aleph_0$. Let $\mathbb{Q}_-$ denote the negative rational numbers, and note also that $\mathbb{Q} = \mathbb{Q}_+ \cup \{0\} \cup \mathbb{Q}_-$ and the union of three countable sets is countable (see Exercise 10 at the end of Section 1.1). Thus $\mathbb{Q}$ is countable.

**Remark 3.** Given a nonempty set $X$, consider $\mathcal{P}(X) = \{S \mid \forall S,\ S \subset X\}$ the *power set* of $X$. It is a good exercise to show that $\mathcal{P}(X)$ is in one-to-one correspondence with the set of all functions from $X$ to $\{0, 1\}$. A function $f : X \to \{0, 1\}$ determines a subset of $X$ by thinking of 0 as meaning "do not include" and 1 as meaning "include." For example, if $X = \{1, 2, 3\}$, the function $f : X \to \{0, 1\}$ defined by $f(1) = 0$, $f(2) = 1$, and $f(3) = 1$ determines the subset $S = \{2, 3\}$. In particular, for each element $x \in X$, we have two choices when making a subset, either include $x$ or do not, and so the cardinality of $\mathcal{P}(X)$ is $2^{|X|}$. For finite sets, it is clear that $|X| < 2^{|X|}$, but what about infinite sets? This question is answered by showing that there is no bijection between $X$ and $\mathcal{P}(X)$ (for proof of this fact consult [43], p. 35.) Hence $|X| < |\mathcal{P}(X)|$.

Surprisingly, "infinity" is not one single number; rather, there are infinitely many different infinite numbers, just like there are infinitely many different finite

numbers. As we defined before sets that correspond bijectively with the natural numbers $\mathbb{N}$ (or the integers $\mathbb{Z}$ or the rational numbers $\mathbb{Q}$) are called *countably infinite*; larger sets are called *uncountable*.

One of the most important sets of numbers is the set $\mathbb{R}$ of real numbers. In the coming sections we study the properties of $\mathbb{R}$ in detail. There are two things to emphasize about $\mathbb{R}$ at this point. The first one is the fact that the reals are uncountable, and the second one is that the cardinality of the real numbers is $c = 2^{\aleph_0}$. Note that some real numbers, precisely all those numbers with finite decimal expression, have two different expansions, one ending in an infinite string of zeros and the other ending with infinite string of nines. For example, $0.125 = 0.1249999\ldots$. For our discussion in Theorem 2 below, we consider the real numbers to be the set of all terminating or infinite decimals with the convention that no decimal expansion can terminate in all nines. Then the decimal expression of a real number is unique, since it does not end all in nines.

**Theorem 2.** *The set of real numbers between* 0 *and* 1 *is not countable.*

*Proof.* This proof is done by contradiction. Assume that there is a bijection between $\mathbb{N}$ and $(0,1)$. Thus for each $n \in \mathbb{N}$, $f(n)$ is a real number in $(0,1)$, and we can represent it using decimal notation as

$$f(n) = 0.a_{n1}a_{n2}a_{n3}a_{n4}\ldots$$

If we write down the $1-1$ correspondence between $\mathbb{N}$ and $(0,1)$ for $n = 1, 2, 3, 4, \ldots$, we get the following array:

$$\begin{array}{rcl} 1 \leftrightarrow f(1) & = & 0.\boldsymbol{a_{11}}a_{12}a_{13}a_{14}\ldots \\ 2 \leftrightarrow f(2) & = & 0.a_{21}\boldsymbol{a_{22}}a_{23}a_{24}\ldots \\ 3 \leftrightarrow f(3) & = & 0.a_{31}a_{32}\boldsymbol{a_{33}}a_{34}\ldots \\ 4 \leftrightarrow f(4) & = & 0.a_{41}\ a_{42}\ a_{43}\ \boldsymbol{a_{44}}\ \ldots \\ \vdots\ \ \vdots & & \vdots\ \ \vdots\ \ \vdots\ \ \vdots \end{array}$$

We are assuming here that every real number you can think of in $(0,1)$ appears somewhere on the above list. Now we are going to define a real number $x \in (0,1)$ which is not in the above list. Let $x = 0.b_1b_2b_3\ldots$. To choose the digit $b_1$, look at the digit $a_{11}$ in the upper left-hand corner of the above array and choose $b_1 \neq a_{11}$. For example, if $a_{11} = 3$, then choose $b_1 = 2$, and if $a_{11} = 2$, then choose $b_1 = 3$. Notice that the real number $x = 0.b_1b_2b_3\ldots$ cannot be equal to $f(1)$. Similarly choose $b_2 \neq a_{22}$, and thus $x \neq f(2)$. Continuing this process we see that $x \neq f(n)$ for any $n \in \mathbb{N}$. In other words the decimal $x = 0.b_1b_2b_3\ldots$ is a real number, and it does not end in all nines but cannot be in the above list since it differs from each number we list in at least one digit. So we reach a contradiction. Consequently, the real numbers between 0 and 1 are not countable. □

**Remark 4.** Since $(0,1)$ is uncountable if and only if $\mathbb{R}$ is uncountable (why? see Exercise 11), the above theorem also shows $\mathbb{R}$ is uncountable. Cantor initially published his discovery that $\mathbb{R}$ is uncountable and later offered this same fact

with an amazing simple proof technique called Cantor's diagonalization method as discussed in the proof of Theorem 2 above. Note that $|\mathbb{N}| = \aleph_0$ and

$$c = |(0,1)| = |\mathbb{R}|,$$

from the inequality $\aleph_0 < c$, the question Cantor asked was does there exists a set $A \subseteq \mathbb{R}$ such that $\aleph_0 < |A| < c$? He conjectured $c$ was an immediate successor of $\aleph_0$; this is called Cantor's "continuum hypothesis." The answer to this question was two unexpected results. In 1940 K. Gödel proved that there was no way to disprove the continuum hypothesis, and in 1963 P. Cohen showed that it was impossible to prove Cantor's conjecture. Putting these two results together tells us that the continuum hypothesis is undecidable. For more information consult [16].

## Equivalence Relations

Sometimes in mathematics, we find that we need a customized mathematical structure with certain desired properties. In other cases, we find that the things we are studying have too much information, with distracting details making it difficult to see what is important and what is irrelevant. In these and other situations, structures known as *relations on a set* can make the difference.

Let $X$ be a set. As we have seen, a function $f : X \to X$ can be understood as a subset of the Cartesian product $X^2$ with the property that every element of $X$ appears exactly once in the first component, e.g., the function $f : \{1,2,3\} \to \{1,2,3\}$ defined by $f(1) = 3$, $f(2) = 1$ and $f(3) = 3$ is really the subset

$$\{(1,3),(2,1),(3,3)\} \subset \{1,2,3\} \times \{1,2,3\}.$$

What about other subsets? More generally, any subset $R \subset X^2$ of the Cartesian product of a set with itself defines a *relation* on $X$. We can think of a relation as a way of comparing two elements of $X$; we will write $xRy$ if the ordered pair $(x,y)$ belongs to $R$.

A relation $\sim$ on $X$ is an *equivalence relation* if $\sim$ satisfies the following three properties:

- For all $x \in X$, we have $x \sim x$ (this is called the *reflexive* property).
- For all $x, y \in X$, $x \sim y \Rightarrow y \sim x$ (this is called the *symmetric* property).
- For all $x, y, z \in X$, $x \sim y$ and $y \sim z$ imply $x \sim z$ (this is called the *transitive* property).

If $\sim$ is an equivalence relation on $X$ and $x \sim y$, we say that $x$ is *equivalent* to $y$.

**Example 8.** The most familiar example of an equivalence relation is *equality*, generally denoted by "=." We can easily verify the properties:

- $x = x$, so $x$ is reflexive.
- If $x = y$, then $y = x$, so $=$ is symmetric.

- If $x = y$ and $y = z$, then $x = z$, so $=$ is transitive.

As a subset of $X \times X$, the relation $=$ consists of ordered pairs where both elements of the pair are the same, i.e.,

$$\{(x, x) \mid x \in X\}.$$

**Example 9.** Let $V$ be a real vector space, and say $\vec{u} \sim \vec{v}$ if there is a nonzero scalar $\alpha \in \mathbb{R}$ so that $\vec{u} = \alpha\vec{v}$. Then $\sim$ is an equivalence relation:

- For any $\vec{u} \in V$, $\vec{u} = 1\vec{u}$, so $\vec{u} \sim \vec{u}$ and $\sim$ is reflexive.
- For any $\vec{u}, \vec{v} \in V$, if $\vec{u} \sim \vec{v}$, then $\vec{u} = \alpha\vec{v}$ for some $\alpha \neq 0$, so $\vec{v} = \alpha^{-1}\vec{u}$ and $\vec{v} \sim \vec{u}$. Thus, $\sim$ is symmetric.
- For any $\vec{u}, \vec{v}, \vec{w} \in V$, if $\vec{u} \sim \vec{v}$ and $\vec{v} \sim \vec{w}$, then we have $\vec{u} = \alpha\vec{v}$ and $\vec{v} = \beta\vec{w}$ for $\alpha, \beta \neq 0$, so $\vec{u} = \alpha\beta\vec{w}$ and $\vec{u} \sim \vec{w}$.

**Example 10.** Let $X = \mathbb{Z} \times \mathbb{Z}$, the set of ordered pairs of integers, and say $(x, y) \sim (x', y')$ if there are nonzero integers $m$ and $n$ such that $(mx, my) = (nx', ny')$. Let us show that $\sim$ is an equivalence relation:

- $(1x, 1y) = (1x, 1y)$, so $\sim$ is reflexive.
- If $(x, y) \sim (x', y')$, then we have $(mx, my) = (nx', ny')$, so $(nx', ny') = (mx, my)$ and $(x', y') \sim (x, y)$; hence, $\sim$ is symmetric.
- If $(x, y) \sim (x', y')$ and $(x', y') \sim (x'', y'')$, then there are integers $n, m$ and $j, k$ such that

  $$(mx, my) = (nx', ny') \quad \text{and} \quad (jx', jy') = (kx'', ky'').$$

  Then we have

  $$(mjx, mjy) = (njx', njy') = (nkx'', nky''),$$

  and $\sim$ is transitive.

**Example 11.** Let $X = \mathbb{Z}$ and fix a positive integer $n$. Say $x \sim y$ if $x - y = nm$ for some integer $m$. Then $\sim$ is an equivalence relation:

- $x - x = 0 = 0n$ so $\sim$ is reflexive.
- If $x \sim y$, then $x - y = nm$ for some $m$; therefore $y - x = -nm = n(-m)$, so $y \sim x$ and $\sim$ is symmetric.
- If $x \sim y$ and $y \sim z$, then $x - y = nm$ and $y - z = nk$; therefore

  $$x - z = x - y + y - z = nm + nk = n(m + k),$$

  and $\sim$ is transitive.

This equivalence relation is known as *congruence modulo* $n$, often denoted by $x \equiv y \ (n)$.

**Example 12.** Let $X = \mathbb{R}^2$, and define $\vec{u} \sim \vec{v}$ if $\vec{u} - \vec{v} \in \mathbb{Z}^2$, i.e., two vectors in the plane are equivalent if their components differ by integers. Then $\sim$ is an equivalence relation; see Exercise 13 below.

Given $x \in X$, the *equivalence class* of $x$, denoted $[x]$, is the set of all elements of $X$ which are equivalent to $x$:

$$[x] = \{y \in X \mid x \sim y\}.$$

We can notice a few interesting things right away. First, equivalence classes are disjoint: Given $x \neq y \in X$, we have either $[x] = [y]$ or $[x] \cap [y] = \emptyset$, since if $z \in [x] \cap [y]$, then $x \sim z$, $y \sim z$, and transitivity says $x \sim y$. In particular, an equivalence relation $\sim$ partitions $X$ into disjoint sets $[x]$. The set of these equivalence classes is called the *quotient space* of $X$ *modulo* $\sim$, written

$$X/_\sim = \{[x] \mid x \in X\}.$$

For example, setting $n = 2$ in Example 11, we have an equivalence relation on $\mathbb{Z}$; $x \sim y \iff x - y = 2m$ for some $m \in \mathbb{Z}$. What are the equivalence classes? Well, for any number $y \in \mathbb{Z}$, we can find its equivalence class

$$[y] = \{x \in X \mid x \sim y\}.$$

For instance,

$$[0] = \{y \in \mathbb{Z} \mid y - 0 = 2m\} = \{y \in \mathbb{Z} \mid y = 2m\},$$

so in fact the class of 0 is all even integers. Moreover

$$[1] = \{y \in \mathbb{Z} \mid y - 1 = 2m\} = \{y \in \mathbb{Z} \mid y = 2m + 1\},$$

and the class of 1 is the set of all odd integers. Thus, the equivalence relation $x \sim y \iff x - y = 2m$ partitions the integers into the sets of even and odd integers.

An equivalence relation on $X$ determines a partition of $X$ into disjoint subsets. It turns out that it works the other way too: If we start with a division of $X$ into disjoint subsets $X = X_1 \cup \cdots \cup X_n$, we can define $x \sim y$ if $x$ and $y$ are in the same $X_k$. We claim that this defines an equivalence relation:

- Clearly, an element $x$ is in the same $X_k$ as itself, so $\sim$ is Reflexive.
- If $x$ and $y$ are in the same $X_k$, then switching their order does not change that they are in the same $X_k$, so $x \sim y$ implies $y \sim x$ and $\sim$ is symmetric.
- If $x$ and $y$ are in the same $X_k$ and $y$ and $z$ are in the same $X_j$, then the fact that the $X_k$'s are disjoint means $y$ is in only one subset, and $x$ and $z$ must both belong to the same subset. Thus, $\sim$ is transitive.

Hence, partitions and equivalence relations are really the same thing.

Third, there is a map $f : X \to X/_\sim$ called the *projection* map defined by

$$f(x) = [x],$$

i.e., sending $x$ to its equivalence class. Given any set $X$ and equivalence relation $\sim$, this projection map is surjective. This also works in the other direction: Given any surjective map $f : X \to Y$, we can think of $f$ as a projection map and identify $Y$ with $X/_\sim$, where $x \sim x'$ means $f(x) = f(x')$. Thus, equivalence relations, partitions, and surjective maps are all really the same thing.

**Example 13.** Many important objects in mathematics arise naturally as quotient sets. For instance, consider the equivalence relation in Example 12. What is the quotient set? Points in the quotient set are equivalence classes in $X$, so to identify the quotient set we need to keep only one point from each equivalence class. Then every point in the plane is equivalent to a point in the unit square $[0, 1] \times [0, 1]$, so we only need the square. But points along the top edge are equivalent to points on the bottom edge, so we can picture gluing the top edge to the bottom edge to obtain a cylinder. The same holds for the sides, so gluing the sides together, we obtain a *torus*, the doughnut as in Figure 1.10.

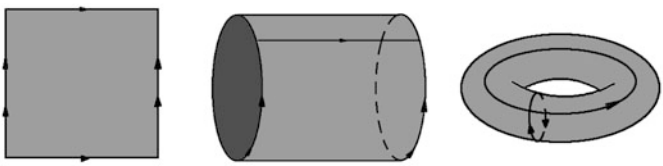


Figure 1.10: Torus as a quotient set

The quotient sets in Examples 10 and 11 are $\mathbb{Q}$ and $\mathbb{Z}_n$, where $\mathbb{Z}_n$ denotes the integers modulo $n$. We will give the definition of $\mathbb{Z}_n$ in Section 1.2 when we explain modular arithmetic. The real numbers, as we will see, can be understood as equivalence classes of certain infinite sequences of rational numbers known as *Cauchy* sequences where two sequences are equivalent if the difference between them approaches zero as $n$ gets large. Even the natural numbers can be understood as equivalence classes of finite sets where two sets are equivalent if there is a bijection between them.

## Axiom of Choice

The *axiom of choice* is one of the several rules that one can use for building sets out of other sets. Roughly speaking, the axiom of choice says that we are allowed to make an arbitrary number of unspecified choices to form a set. It explicitly or implicitly appears in many proofs in mathematics. For example, let us examine the well-known proof that the countable union of countable sets is

countable. The fact that it is a countable union means we are allowed to write out sets in a list

$$X_1, X_1, X_3, \ldots$$

and then the fact that each $X_n$ is countable allow us to list its elements as $x_{n1}, x_{n2}, x_{n3}, \ldots$. We then finish the proof showing some systematic way of counting through $x_{nm}$. In this proof we made an infinite number of unspecified choices; we "choose" a list of elements of $X_n$ without specifying the choice we had made. Moreover since we know nothing about the sets $X_n$, it is impossible to say how we list them. For a long time mathematicians discussed among themselves about the use of the axiom of choice and seemingly contradictory results it produced. Why do people make a fuss about the axiom of choice? The main reason is that if it is used in a proof, then the part of that proof is nonconstructive. Now most seem to agree that the advantages of the axiom of choice outweigh the disadvantages.

**Definition 5.** Suppose $X$ is a set, $I$ is an index set (not assumed to be countable), and $\{X_i\}_{i\in I}$ is a collection of nonempty subsets of $X$. We call $\psi$ a *choice function* if $\psi : I \to \bigcup_{i\in I} X_i$, defined as $i \mapsto x_i$ such that $\psi(i) \in X_i$ for all $i$. The axiom of choice can be stated as follows:

**The Axiom of Choice:** For every collection of nonempty sets, there exists a choice function.

Note that the axiom of choice is a nonconstructive assertion of existence. It postulates the existence of certain objects without giving any indication of how to find these objects. There are various logically equivalent statements to the axiom of choice, such as the Hausdorff Maximality Principle, Zorn's lemma, and Well-Ordering Principle which we state below. Among these are two forms of the axiom of choice that are more often used in mathematics than the basic form we stated above. One is the *well-ordering principle* which states every set can be well-ordered. The other is *Zorn's lemma* which states under certain circumstances "maximal" element exists. This is hugely important; for example, a basis for a vector space is precisely a maximal linearly independent set, and it turns out that Zorn's lemma applies to collections of linearly independent sets in a vector space, which shows that every vector space has a basis.

To define these equivalent concepts, we first need the definitions of *partially ordered set*, *totally ordered set*, and *well-ordered set*.

**Definition 6.** A *partially ordered set* is a set $X$ with a relation $\leq$ that is reflexive, transitive, and antisymmetric. By antisymmetric we mean if $x \leq y$ and $y \leq x$, then $x = y$. In a partially ordered set, an element $y$ is *maximal* if $x \geq y$ implies $x = y$. Also, in a partially ordered set $X$ with $Y \subset X$, an *upper bound* for $Y$ is an element $x \in X$ such that $y \leq x$ for all $y \in Y$.
A *totally ordered set* is a partially ordered set with an additional property that

for any two elements, say $x, y \in X$ ($x$ and $y$ distinct), either $x \leq y$ or $y \leq x$ (but not both). Finally, a *well-ordered set* $X$ is a totally ordered set in which any nonempty subset $E \subset X$ has a least element. That is an element $x^* \in E$ such that $x^* \leq x$ for any other $x \in E$.

**Example 14.** Simple examples of totally ordered sets are $(\mathbb{N}, \leq)$, $(\mathbb{Q}, \leq)$, and $(\mathbb{R}, \leq)$. Let $X$ be a set, and let $\mathcal{P}(X)$ denote the collection of all subsets of $X$. Then it is not hard to see that $(\mathcal{P}(X), \subseteq)$ is a partially ordered set.

**Hausdorff Maximality Principle:** Every partially ordered set $X$ contains a totally ordered subset $Y$ that is maximal with respect to the ordering on $\mathcal{P}(X)$.

**Zorn's Lemma:** If a nonempty partially ordered set has the property that every nonempty totally ordered subset has an upper bound, then the partially ordered set has a maximal element.

**Well-Ordering Principle:** Any set $X$ can be well-ordered.

Perhaps it is obvious that the Well-Ordering Principle implies the Axiom of Choice, because if we well-order $X$, we can choose $x_i$ to be the smallest element of $X_i$, and in this way we have constructed the required choice function. However, it is not so easy to show that the Axiom of Choice implies the Well-Ordering Principle. The proofs of the equivalence of the Axiom of Choice, the Hausdorff Maximality Principle, Zorn's lemma, and the Well-Ordering Principle are all explained in the book by Paul J. Sally [43], p. 40.
We will need the axiom of choice in Sections 2.7 and 2.8 of Chapter 2, when we are discussing non-measurable sets and the proof of the Banach–Tarski paradox. The Banach–Tarski paradox states that there is a way of dividing a solid unit sphere into a finite number of subsets and then reassembling these subsets using rotations, reflections, and translations to form two solid unit spheres. The proof does not provide an explicit way of defining subsets. Why does it seem strange and paradoxical? It is because we feel volume has not been preserved. But how do we know we can give a sensible definition of volume for all subsets of the sphere? This leads us to measure theory and non-measurable sets.

## Exercises

1. For two sets $A$ and $B$, show that the following statements are equivalent:

   a) $A \subseteq B$.

   b) $A \cup B = B$.

   c) $A \cap B = A$.

2. Establish the following set theoretic relations:

a) $A \cup B = B \cup A$ and $A \cap B = B \cap A$ (Commutativity).

b) $A \cup (B \cup C) = (A \cup B) \cup C$ and $A \cap (B \cap C) = (A \cap B) \cap C$ (Associativity).

c) $A \cup (B \cap C) = (A \cup B) \cap (B \cup C)$ and $A \cap (B \cup C) = (A \cap B) \cup (B \cap C)$ (Distributivity).

d) $A \subseteq B \iff B^c \subseteq A^c$.

e) $A \setminus B = A \cap B^c$.

f) $(A \cup B)^c = A^c \cap B^c$ and $(A \cap B)^c = A^c \cup B^c$. (De Morgan's laws).

3. If $A, B$, and $C$ are sets, show that

a) $A \times B = \emptyset \iff A = \emptyset$ or $B = \emptyset$.

b) $(A \cup B) \times C = (A \times C) \cup (B \times C)$.

c) $(A \cap B) \times C = (A \times C) \cap (B \times C)$.

4. Suppose $f : A \to B$ and $g : B \to C$ are functions, show that

a) If both $f$ and $g$ are one-to-one, then $g \circ f$ is one-to-one.

b) If both $f$ and $g$ are onto, then $g \circ f$ is onto.

c) If both $f$ and $g$ are bijection, then $g \circ f$ is a bijection.

5. For an arbitrary function $f : X \to Y$, prove that the following relations hold:

a) $f(\bigcup_{i \in I} A_i) = \bigcup_{i \in I} f(A_i)$.

b) $f(\bigcap_{i \in I} A_i) \subseteq \bigcap_{i \in I} f(A_i)$.

c) Give a counterexample to show that $f(\bigcap_{i \in I} A_i) = \bigcap_{i \in I} f(A_i)$ is not always true.

6. Given $f : A \to B$, suppose there exists two functions $g : B \to A$ and $h : B \to A$ such that $f \circ g = I_B$ and $h \circ f = I_A$. Show that $f$ is a bijection and that $f^{-1} = g = h$.

7. For a function $f : X \longrightarrow Y$, show that the following statements are equivalent:

a) $f$ is one-to-one.

b) $f(A \cap B) = f(A) \cap f(B)$ holds for all $A, B \in \mathcal{P}(X)$.

8. Let $A$ be a set, and let $\mathcal{P}(A)$ denote the set of all subsets of $A$ (i.e., the power set of $A$). Prove that $A$ and $\mathcal{P}(A)$ do not have the same cardinality.

9. If $A$ and $B$ are sets, then show that

   a) $\mathcal{P}(A) \cup \mathcal{P}(B) \subseteq \mathcal{P}(A \cup B)$.

   b) $\mathcal{P}(A) \cap \mathcal{P}(B) = \mathcal{P}(A \cap B)$.

10. Prove $(1+x)^n \geq 1 + nx$ for all real $x > -1$ and all positive integers $n$ (*Bernoulli's inequality*).

11. Prove $2^n \geq n^2$ for all $n \geq 4$.

12. An *algebraic number* is a root of a polynomial whose coefficients are rational. Show that the set of all algebraic numbers is countable.

13. Show that a countable union of finite or countable sets is countable.

14. Show that $(0,1)$ is uncountable if and only if $\mathbb{R}$ is uncountable.

15. Let $X = \mathbb{R}^2$, and define $\vec{u} \sim \vec{v}$ if $\vec{u} - \vec{v} \in \mathbb{Z}^2$, i.e., two vectors in the plane are equivalent if their components differ by integers. Then show that $\sim$ is an equivalence relation.

## 1.2 Number Systems

For most of us, our first exposure to mathematical reasoning involves numbers and computation, and sadly too many of us never encounter any mathematics past this early stage. Even worse, the interesting aspects of numbers are too often ignored in favor of mechanical algorithms for computation. In this section we will explore the various number systems that are used throughout mathematics. We will focus on the interesting side too often ignored in favor of the mechanics of computation, which we assume the reader is familiar with from high school algebra.

### The Natural Numbers

Recalling the definition given in Example 7, the *natural numbers*

$$\mathbb{N} = \{1, 2, 3, \ldots, n, \ldots\}$$

are the kinds of numbers, which can be cardinalities of finite sets. Indeed, this is how we first conceptualize natural numbers – to teach a child the concept of the number two, we might show him her many examples of pairs of things.

Thinking of the natural numbers as cardinalities of finite sets gives us two operations on $\mathbb{N}$: *addition* comes from taking unions of disjoint sets, and *multiplication* comes from taking Cartesian products:

$$|A| + |B| = |A \cup B| \quad (\text{if } A \cap B = \emptyset) \quad \text{and} \quad |A| \times |B| = |A \times B|.$$

In order to minimize writing, we generally omit the "$\times$" symbol when multiplying, writing "$xy$" instead of "$x \times y$ ."

These operations of addition and multiplication have some important properties:

- **Commutativity.** If we are adding or multiplying two natural numbers, the order of the two numbers does not matter:

$$x + y = y + x \quad \text{and} \quad xy = yx.$$

- **Associativity.** If we are adding or multiplying three numbers, it does not matter what order we resolve the operations:

$$x + (y + z) = (x + y) + z \quad \text{and} \quad x(yz) = (xy)z.$$

- **Identity Elements.** There are natural numbers 0 and 1 which are *identity elements* with respect to addition and multiplication, respectively:

$$0 + x = x \quad \text{and} \quad 1x = x.$$

Not every operation on a set has the properties above; they are useful enough that mathematicians have a name for a set with an operation satisfying all the properties above. A set with an associative operation is called a *semigroup*; a semigroup with an identity element is called a *monoid*, and if the operation is commutative, we have a *commutative semigroup* or *commutative monoid*.

## The Integers and Rational Numbers

As someone playing with sets quickly observes, some additions and multiplications can be undone, while others cannot – we can undo multiplication by two if our input is even, but not if it is odd.

One approach is to define new operations that undo the old ones, e.g., subtraction and division. However, there are some problems (or at least, inelegancies) in this approach – the "undoing operations" may not share the nice properties of the operation they undo, and worse, they may not even make sense at all in some cases.

For example, subtraction and division are both non-commutative and non-associative:

$$3 - 2 \neq 2 - 3 \quad \text{and} \quad 2 \div 3 \neq 3 \div 2,$$

while

$$1 - (2 - 3) = 1 - (-1) = 2 \neq -4 = -1 - 3 = (1 - 2) - 3$$

and

$$1 \div (2 \div 3) = \frac{3}{2} \neq \frac{1}{6} = (1 \div 2) \div 3.$$

More problematic is the fact that the operation may not even be defined for every pair of natural numbers. For example, division by zero does not make any sense – if $x \div 0 = y$, then $0y = x$, which is not possible if $x$ and $y$ are both nonzero natural numbers.

A better solution is to expand the set of numbers we are using so that we can undo addition with another addition. That is, for every natural number $x \neq 0$, we would like there to be another number $-x$ such that $-x + x = 0$ and a number $x^{-1}$ such that $xx^{-1} = 1$. Including these gives us the *negative integers* and *rational numbers*, respectively. In particular, subtraction $x - y$ is simply addition to $x$ of $-y$ and division $x \div y$ is simply multiplication of $x$ by $y^{-1}$. The numbers $-x$ and $x^{-1}$ are the *additive inverse* and *multiplicative inverse*, respectively, of $x$.

The rational numbers are the first example we usually encounter of an quotient set. For example, "$\frac{1}{2}$" and "$\frac{2}{4}$" are different literal fractions – they have different numerators and different denominators, so they are not identical – but they represent the same ratio. A rational number is not a single fraction, but an infinite set of fractions each of which can serve as a representative of its set. Indeed, to add fractions it is necessary to select representatives with the same denominator.

Negative numbers are often initially viewed with suspicion – after all, how can anything be less than zero? Of course, in the modern world, all it takes is having a credit card to know that negative numbers are perfectly real. We can also think of negative numbers as telling us direction – a speed of $-5$ miles per hour indicates that the car is moving at 5 mph in reverse.

The standard symbols for the sets of integers and rational numbers are $\mathbb{Z}$ and $\mathbb{Q}$. The "$\mathbb{Q}$" stands for "quotient," but the choice of "$\mathbb{Z}$" for "integers" seems strange – should not it be "$\mathbb{I}$" for "integer?" In fact, the symbol *is* chosen for the first letter of the word for integers; it is just that it was named in German, with "$\mathbb{Z}$" short for *Zahlen*, the German word for "integer." The rational numbers contain the integers as a subset which in turn contain the natural numbers as a subset. It is natural to ask the question what else do we know for $\mathbb{Z}$? We have inequalities. The sign "$<$" is a relation on $\mathbb{Z}$ that is subject to the following *order properties*:

$O_1$: If $a, b \in \mathbb{Z}$, then one and only one of the following holds: $a < b$, $a = b$, or $b < a$ (trichotomy).

$O_2$: If $a, b, c \in \mathbb{Z}$ with $a < b$ and $b < c$, then $a < c$ (transitivity).

$O_3$: If $a, b, c \in \mathbb{Z}$ with $a < b$, then $a + c < b + c$ (addition).

$O_4$: If $a, b, c \in \mathbb{Z}$ with $a < b$ and $0 < c$, then $ac < bc$ (multiplication by positive elements).

We adopt the standard notation. That is, if $a < b$, we say that $a$ is less than $b$. If $a < b$ or $a = b$, we say that $a$ is less than or equal to $b$ and write $a \leq b$. Instead of $a < b$, we may write $b > a$ and say $b$ is greater than $a$. The notation "$\geq$" is defined analogously.

## Fields

The rational numbers are our first example of an important mathematical structure known as a *field*. A field is a set $\mathbb{F}$ with addition and multiplication operations such that

- Both operations are associative and commutative.
- The operations interact via the *distributive law*

$$a(b+c) = ab + ac \quad \text{and} \quad (a+b)c = ac + bc.$$

- There are additive and multiplicative identity elements (i.e., 0 and 1).
- Every element $x \in \mathbb{F}$ has an additive inverse $-x \in \mathbb{F}$ and every nonzero element $x \in \mathbb{F}$ has a multiplicative inverse $x^{-1} \in \mathbb{F}$.

We can use the field properties to prove theorems that hold for all fields. For example, see the following theorem:

**Theorem 3** (Additive Cancellation Law)**.** *If $\mathbb{F}$ is a field and $\alpha, \beta, \gamma \in \mathbb{F}$ satisfy $\alpha + \beta = \alpha + \gamma$, then $\beta = \gamma$.*

*Proof.* Consider $-\alpha + (\alpha + \beta)$. Since $\alpha + \beta = \alpha + \gamma$, we have

$$-\alpha + (\alpha + \beta) = -\alpha + (\alpha + \gamma).$$

Now, the fact that the addition is associative says

$$(-\alpha + \alpha) + \beta = (-\alpha + \alpha) + \gamma,$$

and we then have

$$0 + \beta = 0 + \gamma.$$

This says

$$\beta = \gamma$$

as required. □

Once we have proved a result, we can use it to prove other results. For example, see the following theorem:

**Theorem 4.** *For any element $\alpha$ in a field $\mathbb{F}$, $0\alpha = 0$.*

*Proof.* First, we note that $0 = 0 + 0$. Then

$$0\alpha = (0 + 0)\alpha = 0\alpha + 0\alpha$$

by the distributive law; on the other hand

$$0\alpha = 0\alpha + 0 = 0\alpha + 0\alpha.$$

Then by additive cancellation we have $0 = 0\alpha$. □

**Theorem 5.** *If $\mathbb{F}$ is a field in which 0 has a multiplicative inverse, then every element of $\mathbb{F}$ equals zero.*

*Proof.* Let $\mathbb{F}$ be a field in which 0 has a multiplicative inverse $0^{-1}$. Then for any $\alpha \in \mathbb{F}$, we have

$$\alpha = 1\alpha = (00^{-1})\alpha = 0(0^{-1}\alpha) = 0.$$ □

Note that this explains why we cannot just declare $\infty = \frac{1}{0}$ and make math much simpler – any field in which zero has a multiplicative inverse contains only one element, namely zero. Thus, we have to either give up the field axioms, give up on zero having a multiplicative inverse, or settle for a number system with only one number.

## Modular Arithmetic

Another very useful number system (a well, infinite set of number systems) which is unfamiliar to most outside of mathematics and computer science is the *integers modulo n*. Let $n$ be a positive integer. Then the set of integers modulo $n$ is the set

$$\mathbb{Z}_n = \{0, 1, 2, \ldots, n - 1\}$$

of remainders after long division by $n$. An element of $\mathbb{Z}_n$ is an equivalence class, like rational numbers, except that where two fractions represent the same rational number if one can be obtained from the other by multiplying top and bottom by the same integer, and two integers represent the same element of $\mathbb{Z}_n$ if one can be obtained from the other by adding a multiple of $n$.

In particular, we can do arithmetic mod $n$ by doing ordinary arithmetic with the extra rule that whenever our numbers get outside the range $\{0, 1, 2, \ldots, n - 1\}$, we divide by $n$ and keep only the remainder, i.e., add or subtract $n$ repeatedly until we are back in $\{0, 1, 2, \ldots, n-1\}$. Where $\mathbb{Z}$ forms a number line, $\mathbb{Z}_n$ forms a number circle, like an analog clock – in fact, we use mod 12 (and mod 60) arithmetic for telling time (Figure 1.11).

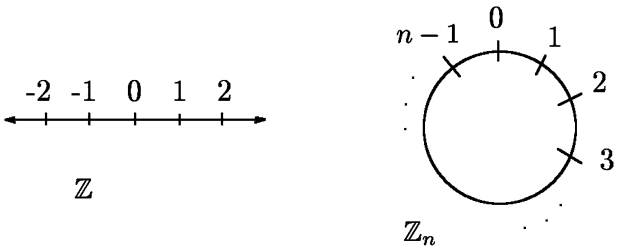



Figure 1.11: Integers and integers modulo $n$

**Example 15.** In $\mathbb{Z}_4$, we have

$$2(3+3)+3(2-3) = 2(6)+3(-1) = 2(2)+3(3) = 4+9 = 0+1 = 1.$$

**Example 16.** We can do reduction mod $n$ at every step, or we can save it all up for the last step, and we will get the same result. For example, in $\mathbb{Z}_5$, $(2+4)3-4(3+3) = 2$:

$$\begin{aligned}(2+4)3-4(3+3) &= (6)3-4(6)\\ &= (1)3-4(1)\\ &= 3-4\\ &= -1\\ &= 4\end{aligned}$$

or

$$\begin{aligned}(2+4)3-4(3+3) &= 18-24\\ &= -6\\ &= -1\\ &= 4.\end{aligned}$$

Modular arithmetic differs from ordinary arithmetic in some important ways.

- A *zero divisor* $x \neq 0$ has the property that $xy = 0$ for some $y \neq 0$. In integer arithmetic, there are no zero divisors, but there can be zero divisors in $\mathbb{Z}_n$ depending on $n$. For instance, in $\mathbb{Z}_4$, 2 is a zero divisor since $2(2) = 4 = 0$.
- An element $x \in \mathbb{Z}_n$ is a *unit* if it has a multiplicative inverse, i.e., if there is an $x^{-1} \in \mathbb{Z}_n$ such that $xx^{-1} = 1$. For example, in $\mathbb{Z}_7$, we have

| $x$ | $x^{-1}$ |
|---|---|
| 1 | 1 |
| 2 | 4 |
| 3 | 5 |
| 4 | 2 |
| 5 | 3 |
| 6 | 6 |

since $2(4) = 8 = 1$, $3(5) = 15 = 1$, and $6(6) = 36 = 1 \bmod 7$. In a field, every nonzero element must be a unit.

We have seen that $\mathbb{Q}, \mathbb{R}$, and $\mathbb{C}$ are fields, while $\mathbb{N}$ and $\mathbb{Z}$ are not. $\mathbb{N}$ lacks additive and multiplicative inverses for elements greater than 1, and while $\mathbb{Z}$

includes additive inverses for everything, the only elements in $\mathbb{Z}$ with multiplicative inverses are 1 and $-1$. We say $\mathbb{Z}$ is an *integral domain*, a type of number system which does not contain zero divisors.

**Example 17.** Note that $\mathbb{Z}_5$ is a field. To prove this, one needs to verify that all of the field axioms are satisfied. Most of these are straightforward: Multiplication and addition are associative, commutative, and distributive in $\mathbb{Z}_5$ because they are associative, commutative, and distributive in $\mathbb{Z}$ before reducing mod 5, and reducing equal quantities mod 5 yields equal quantities in $\mathbb{Z}_5$. The interesting part is to check the additive and multiplicative inverses, which we can do by brute force since $\mathbb{Z}_5$ has only five elements.

| $\alpha$ | $-\alpha$ | | $\alpha^{-1}$ | |
|---|---|---|---|---|
| 0 | 0 | $0+0=0$ | $-$ | |
| 1 | 4 | $1+4=5=0$ | 1 | $1(1)=1$ |
| 2 | 3 | $2+3=5=0$ | 3 | $2(3)=6=1$ |
| 3 | 2 | $3+2=5=0$ | 2 | $3(2)=6=1$ |
| 4 | 1 | $4+1=5=0$ | 4 | $4(4)=16=1$ |

Note that $\mathbb{Z}_n$ may or may not be a field depending on the value of $n$; see the exercises for more.

### Exercises

1. Prove that in every field, we have the *multiplicative cancellation property*, i.e., that $\alpha\beta = \alpha\gamma$ implies $\beta = \gamma$.

2. Recall that a nonzero element of a number system $\beta \neq 0$ is a *zero divisor* if there is a nonzero element $\alpha \neq 0$ such that $\alpha\beta = 0$ and that $\beta$ is a *unit* if there exists an element $\beta^{-1}$ such that $\beta\beta^{-1} = 1$. Prove that no element can be both a unit and a zero divisor.

3. Make the addition and multiplication tables for the field $\mathbb{Z}_7$.

4. Identify the units of $\mathbb{Z}_{10}$ by finding their multiplicative inverses and the zero divisors.

5. For which values of $n$ is $\mathbb{Z}_n$ a field? Make a conjecture, test it, and try to prove it.

6. Say that an integer $p$ *divides* an integer $n$ if $n = pq$ for some integer $q$, and say that $p$ is *irreducible* if whenever $p$ divides $nm$ then either $p$ divides $n$ or $p$ divides $m$. Show that $\sqrt{p}$ is irrational for every irreducible integer $p$.

## 1.3 Completeness and the Real Number System

What lies beyond the rational numbers? To answer this question, we begin with an observation.

**Observation 5.** If $n$ and $m$ are odd, then $nm$ is odd.

From this observation, we can draw an almost immediate conclusion: If a number $n^2$ is a perfect square, it is either odd (if $n$ is odd) or divisible by 4 (if $n$ is even, then $n = 2m$ for one integer $m$, so $n^2 = (2m)(2m) = 4m^2$). Then we have the following:

**Corollary 1.** *Every even perfect square is divisible by* 4.

This simple observation has a profound consequence:

**Theorem 6.** *There is no rational number* $\frac{p}{q}$ *(where* $p, q \in \mathbb{Z}$ *and* $q \neq 0$*) whose square is 2. That is,* $\sqrt{2}$ *is irrational.*

The argument we will use to prove Theorem 6 is called *proof by contradiction*. The strategy is to assume that there is a rational number whose square is 2 and then show that this assumption leads to a self-contradiction.

*Proof.* Suppose $\sqrt{2} = \frac{p}{q}$, where $p$ and $q$ are both integers in the lowest terms, i.e., with no common divisor greater than 1 (since if there were a common divisor, we could cancel it). Then $2 = \frac{p^2}{q^2}$, so $p^2 = 2q^2$ and $p^2$ is even. Then $p^2$ is an even perfect square, and hence by Corollary 1, $p^2$ is divisible by 4. Then $p^2 = 2q^2 = 4x$ for some integer $x$; then in particular, $q^2 = 2x$. So $q^2$ is even, which says $q$ is even. But then $p$ and $q$ are both even, a contradiction since we chose $\frac{p}{q}$ to be in the lowest terms. That is, *it is impossible* to choose integers $p$ and $q$ with the greatest common divisor 1 so that $\sqrt{2} = \frac{p}{q}$. Thus, there is no rational number $\frac{p}{q}$ whose square is 2. □

Hence, we see that there are numbers which are not rational, like $\sqrt{2}$. A number $x_0$ in $\mathbb{R}$ is called *algebraic* if it satisfies the equation

$$P(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n = 0$$

for some polynomial $P$ whose coefficients $a_0, a_1, \ldots, a_n$ are integers, not all zero. Every rational number is algebraic and the number $\sqrt{2}$ is also algebraic, since it satisfies the equation $x^2 - 2 = 0$. Real numbers that are not algebraic are called *transcendental*. It is known, for example, that the numbers $\pi$ and $e$ are transcendental. For more on transcendental numbers, see Chapter 2 in [20]. In the following we will discuss the properties of real numbers in more detail.

## The Real Numbers

The set of *real numbers*, denoted $\mathbb{R}$, contains the rational numbers $\mathbb{Q}$ as a subset. The operations of addition and multiplication on $\mathbb{Q}$ extend to all of $\mathbb{R}$ in such a

way that every element of $\mathbb{R}$ has an additive inverse and every nonzero element of $\mathbb{R}$ has a multiplicative inverse. Real numbers also obey the *order axioms* from $O_1$ through $O_4$ given for $\mathbb{Q}$. In fact $\mathbb{R}$ is an *ordered field*, which contains $\mathbb{Q}$ as a subfield. We will look at the real number system in much more detail later, when we define the concept of "least upper bound."

It can be shown that the real numbers are the only ordered field with the property known as *completeness*. Geometrically, the real numbers form a continuous line with no gaps or holes. Although field axioms are fundamental to the real numbers, by themselves they do not characterize $\mathbb{R}$. In other words, the axioms of an ordered field are not enough to characterize $\mathbb{R}$ uniquely. For example, the set of rational numbers $\mathbb{Q}$ is a subfield of $\mathbb{R}$ and also obeys these axioms. The one additional axiom that distinguishes $\mathbb{R}$ from $\mathbb{Q}$ is called the *completeness axiom*. The rational numbers by themselves are inadequate for analysis even though there are many rational numbers, in fact, a countably infinite number of them exist. Moreover, given any two rational numbers you can think of say, $\frac{1}{100}$ and $\frac{1}{101}$, one can find infinitely many rational numbers between these two. Yet it seems like there are not just enough of them. We have seen that the simple-looking equation $x^2 = 2$ has no solution in the set of rational numbers $\mathbb{Q}$. We showed $\sqrt{2}$ is a real number, and it is the solution of the equation $x^2 = 2$, but $\sqrt{2} \notin \mathbb{Q}$. Thus the set of rational numbers contains "gaps." We need a concept that clarifies what we mean by $\mathbb{R}$, which does not contain any gaps or holes which also defines the difference between $\mathbb{R}$ and $\mathbb{Q}$.

## Axiom of Completeness

**Axiom of Completeness:** Every nonempty set of real numbers that is bounded above has a least upper bound.

To understand the axiom of completeness, let us first define *bounded sets* and *least upper bound* of a set.

**Definition 7.** A nonempty set $A \subset \mathbb{R}$ is *bounded above* if there exists a number $M$ with $a \leq M$ for all $a \in A$. Then $M$ is called an upper bound for $A$. Similarly the set $A$ is *bounded below* if there exists a number $m$ with $a \geq m$ for all $a \in A$. In this case $m$ is a lower bound for $A$.

Note that a given set can have many upper or lower bounds, for example, if $A = (1, 2)$, then A is bounded below by 1 or by $-4, -36$, etc. $A$ is bounded above by 2, or $3, 4, 100$, etc. But if the set under consideration is $A = (1, \infty)$, then $A$ has lower bounds but has no upper bound.

**Definition 8.** Let $A$ be a nonempty subset of $\mathbb{R}$ that is bounded above. The *supremum* or *least upper bound* of $A$ is a number $s$ such that

a) $s$ is an upper bound of $A$.

b) If $b$ is another upper bound for $A$, then $s \leq b$.

Similarly, let $A$ be a nonempty subset of $\mathbb{R}$ that is bounded below. The *infimum* or *greatest lower bound* of $A$ is a number $s'$ such that

a) $s'$ is a lower bound of $A$.

b) If $c$ is another lower bound for $A$, then $s' \geq c$.

We write $\sup A$ or $\inf A$ to denote the supremum or the infimum of the set $A$, respectively.

Note that the definition of infimum or supremum has two parts. For the least upper bound, the first part asserts that $\sup A$ is an upper bound, and the second part states that $\sup A$ must be the least one. It is not clear that supremum and infimum will always exist. However a finite set will always have a supremum or infimum. We just pick the biggest and the smallest element of the set, respectively. Although a set can have many upper bounds, it can only have one least upper bound. To see this assume the set $A$ has two least upper bounds, say $s_1$ and $s_2$. Then from the definition of the least upper bound, we can assert $s_1 \leq s_2$ and $s_2 \leq s_1$; thus, $s_1 = s_2$ and the least upper bounds are unique (Figure 1.12).

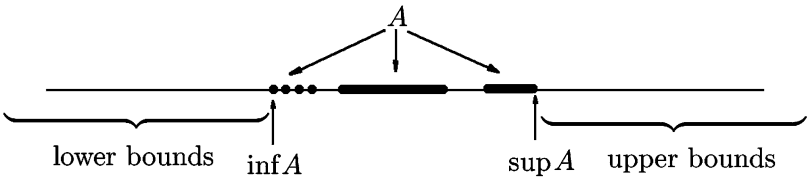



Figure 1.12: Sup A and inf A

**Example 18.**

a) Let $A = \{-1, 3, 8, 10\}$; then $\inf A = -1$ and $\sup A = 10$. Both the infimum and the supremum belong to $A$.

b) Let $B = \{1, 3, 5, 7, \dots\}$. Then $\inf B = 1$, but $\sup B$ does not exist.

c) Let $C = \left\{ \dfrac{2\pi}{n} : \quad n \in \mathbb{N} \right\}$. Then $\sup C = 2\pi$ and $\inf C = 0$. The infimum does not belong to the set $C$.

**Remark 5.** Let $A$ and $B$ be nonempty bounded subsets of $\mathbb{R}$; if

$$A + B := \{a + b : a \in A \text{ and } b \in B\},$$

then it can be shown that $\sup(A + B) = \sup(A) + \sup(B)$.

**Example 19.** Let $A \subset \mathbb{Q}$ be defined as

$$A = \{r \in \mathbb{Q} : \quad r^2 < 2\},$$

and suppose we are looking for the least upper bound for $A$. Since $\sqrt{2} = 1.4142\ldots$, we might guess that $b_1 = \dfrac{142}{100}$ is an upper bound, but it is not the least one; for example, $b_2 = \dfrac{1415}{1000}$ is an upper bound smaller than the first one. Thus the question is can one find the smallest one? In the set of rational numbers we cannot.

We now give a "characterization" condition for the least upper bounds.

**Lemma 1** (Characterization of "sup")**.** *Assume $s \in \mathbb{R}$ is an upper bound for a nonempty set $A \subset \mathbb{R}$. Then $s = supA$ if and only if for every $\epsilon > 0$ there exists an element $x \in A$ such that $s - \epsilon < x$ (Figure 1.13).*

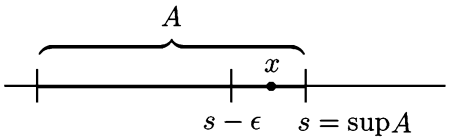



Figure 1.13: Characterization of "sup"

*Proof.* $\Rightarrow$) Assume $s = \sup A$, and $\epsilon > 0$. We must produce an $x \in A$ such that $s - \epsilon < x$. If there is no such $x$, we would have $s \geq x + \epsilon$ for every $x \in A$, that is, $s - \epsilon \geq x$. Clearly, $s - \epsilon$ is an upper bound strictly less than $s$, and therefore $s$ is not the least upper bound, which contradicts our hypothesis.
($\Leftarrow$ suppose $s$ satisfies the given condition. Let $s^*$ be an upper bound of $A$. According to the definition of $\sup A$, we must show $s \leq s^*$. Suppose $s > s^*$, and then if we let $\epsilon = s - s^*$, $s^* = s - \epsilon$ and $s^* \geq x$ for all $x \in A$, which implies $s - \epsilon \geq x$ or $s \geq x + \epsilon$ and so our condition fails. Our assumption $s > s^*$ is wrong, and therefore $s \leq s^*$. Thus we verified both conditions in the definition of the least upper bound. □

**Remark 6.** It is certainly the case that all of the above conclusions about $\sup A$ have analogous versions for $\inf A$. Furthermore the Axiom of Completeness can also be expressed as follows:

Let $A$ be a nonempty set in $\mathbb{R}$ that has a lower bound. Then $A$ has a greatest lower bound.

The first application of the Axiom of Completeness is the Nested Interval Property, which expresses the fact that the real line contains no "gaps." A collection of intervals $I_1, I_2, I_3, \ldots$ is called *nested* if

$$I_1 \supseteq I_2 \supseteq I_3 \supseteq \ldots.$$

If the intervals $I_n$ are open intervals:

$$(0, 1) \supseteq (0, 1/2) \supseteq (0, 1/3) \supseteq \ldots,$$

then even though each interval contains infinitely many points, the intersection $\bigcap_{n=1}^{\infty} I_n = \emptyset$. The intervals $\{[n, \infty)\}_{n=1}^{\infty}$ are closed and nested but not bounded, and their intersection is empty. The story is different if we take both closed and bounded intervals, for example, if we consider the nested collection

$$[0, 1] \supseteq [0, 1/2] \supseteq [0, 1/3] \supseteq \dots,$$

then their intersection has one point $\bigcap_{n=1}^{\infty} I_n = \{0\}$ and thus is not empty.

**Theorem 7** (Nested Intervals Property)**.** *Let $I_1, I_2, \dots$ be a sequence of nonempty closed and bounded intervals which are nested in the sense that each $I_n$ contains $I_{n+1}$ for every $n \in \mathbb{N}$. Then the intersection $\bigcap_{n=1}^{\infty} I_n \neq \emptyset$.*

*Proof.* Let us denote the closed and bounded intervals by $I_n = [a_n, b_n]$ for all positive integers $n$. Then the nesting condition means

$$a_1 \leq a_2 \leq a_3 \leq \dots \leq b_3 \leq b_2 \leq b_1.$$

Now consider the "left hand point set" $A = \{a_1, a_2, a_3, \dots\}$. Because the intervals are nested, notice that every $b_n$ is an upper bound of the set $A$, and thus we can set $x = \sup A$. Now, consider a particular $I_n = [a_n, b_n]$. Since $x$ is an upper bound for $A$, $a_n \leq x$. The fact that each $b_n$ is an upper bound and $x$ is the least upper bound for $A$ implies $x \leq b_n$. Thus we have

$$a_n \leq x \leq b_n \quad \text{for every choice} \quad n \in \mathbb{N},$$

which means $x \in I_n$ for each $n$, therefore $x \in \bigcap_{n=1}^{\infty} I_n$, and the intersection is not empty. □

Note that if the interval's length shrinks to zero, then the intersection is a single point. In particular, if $A = \{a_1, a_2, a_3, \dots\}$ and $B = \{b_1, b_2, b_3, \dots\}$, then by the axiom of completeness $a = \sup A$ and $b = \inf B$ lie in all the $I_n$. If the length of $I_n$ shrinks to zero, then $a = b$.

## The Density of $\mathbb{Q}$ in $\mathbb{R}$

The set of rational numbers $\mathbb{Q}$ contains the set of natural numbers $\mathbb{N}$ and both are contained in $\mathbb{R}$. Now we ask "how do $\mathbb{Q}$ and $\mathbb{N}$ fit inside $\mathbb{R}$?" This will lead us to the density of rationals in the set of real numbers. We say $\mathbb{Q}$ is dense in $\mathbb{R}$ when for every two real numbers $a$ and $b$ with with $a < b$ there exists a rational number $r$ such that $a < r < b$. We now present some facts that follow from the least upper bound property and the properties of integers. The last property is the Archimedean property of the real numbers.

**Theorem 8.**

*a) Given any real number $M$, there exists a positive integer $n$ such that $n > M$ ($\mathbb{N}$ is unbounded).*

*b) For every positive $\epsilon$, no matter how small, there exists a positive integer $n$ with $0 < \dfrac{1}{n} < \epsilon$ (squeeze in).*

*c) Given any real numbers $a$ and $b$ with $0 < a < b$, there exists a positive integer $n$ with $na > b$ (Archimedean property).*

*Proof.* To show $\mathbb{N}$ is unbounded, we argue for a contradiction. Assume that there is no integer $n$ with $n > M$. Then $n \leq M$ for all $n$, which means $M$ is an upper bound for $\mathbb{N}$. Thus by the completeness principle $\mathbb{N}$ has a supremum, say $\lambda = \sup \mathbb{N}$. Since $\lambda$ is the *least* upper bound, $\lambda - 1$ is not an upper bound, and so there is some positive integer $n_1$ for which $n_1 > \lambda - 1$ or equivalently $n_1 + 1 > \lambda$. This is a contradiction to the assumption that no integer exceeds $\lambda$. The second property (squeeze in) follows from the first, and for the Archimedean property, observe that for positive $a$ and $b$

$$na > b \quad \text{if and only if} \quad n > \frac{a}{b},$$

and since $\mathbb{N}$ is unbounded, $n > \dfrac{a}{b}$ must hold for sufficiently large $n$. □

The Archimedean property asserts that given a positive number $a$ as small as one wants, if one takes steps of size $a$, one will cover and exceed a trip of distance $b$. In other words, given a positive number $a$, even if it is very small, its successive multiples $a, 2a, 3a, \ldots, na$ will eventually exceed any proposed bound $b$. The next theorem shows that both rational and irrational numbers sit "tightly" inside $\mathbb{R}$.

**Theorem 9.** *$\mathbb{Q}$ is dense in $\mathbb{R}$. That is, for every two real numbers $a$ and $b$ with $a < b$, there exists a rational number $r$ such that $a < r < b$ (Figure 1.14).*

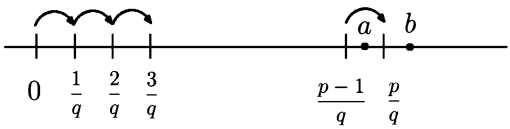



Figure 1.14: Density of $\mathbb{Q}$ in $\mathbb{R}$

*Proof.* Assume $0 \leq a < b$. The case where $a < 0$ follows from this case. Since a rational number is a quotient of two integers, we must find two integers $p$ and $q \neq 0$ so that $r = \dfrac{p}{q} \in (a, b)$. First choose $q$ large enough so that consecutive increments of size $\dfrac{1}{q}$ will eventually be in $(a, b)$. Using the order properties, we

can pick $q \in \mathbb{N}$ so that $\frac{1}{q} < b - a$. Multiplying the inequality $a < \frac{p}{q} < b$ by $q$ gives

$$qa < p < qb.$$

We have already chosen $q$, and now we choose $p$ to be the smallest natural number greater than $qa$. Equivalently, choose $p \in \mathbb{N}$ so that

$$p - 1 \leq qa < p.$$

Now from $qa < p$, it follows that $a < p/q$. On the other hand, $p-1 \leq qa$ implies

$$\begin{aligned} p &\leq qa + 1, \\ &< q(b - 1/q) + 1 \\ &= qb. \end{aligned}$$

Now the inequality $p < qb$ implies $\frac{p}{q} < b$; since we also have $a < \frac{p}{q}$, we proved the desired result. □

**Corollary 2.** *The set of irrational numbers $\mathbb{I} := \mathbb{R} \setminus \mathbb{Q}$ is also dense in $\mathbb{R}$. That is, given any two real numbers $a < b$, there exists an irrational number $\alpha$ such that $a < \alpha < b$.*

Now we introduce a special function, called *the absolute value function*, that is so important that it merits a special notation $|x|$ in place of the usual notation $f(x)$.

**Definition 9.** The absolute value or the magnitude of $x \in \mathbb{R}$ is a function defined as

$$|x| = \begin{cases} x & \text{if } x \geq 0 \\ -x & \text{if } x < 0. \end{cases}$$

The distance between two elements $x$ and $y$ is then $|x - y|$.

Certain properties of the absolute value function are worth singling out both for their own value and because of generalizations to abstract spaces later.

**Lemma 2.** *The absolute value function satisfies the following properties:*

*a)* $|x| \geq 0$ *for every* $x$.

*b)* $|x| = 0$ *if and only if* $x = 0$.

*c)* $|xy| = |x||y|$.

*d)* $|x + y| \leq |x| + |y|$ *(triangle inequality).*

*e)* $||x| - |y|| \leq |x - y|$ *(reverse triangle inequality).*

*Proof.* The proofs of a), b), and c) are straightforward. For the proof of triangle inequality, consider the different cases that arise when $x, y$, and $x+y$ are positive and negative. For example, if $x \geq 0,\ y \leq 0$, and $x+y \geq 0$, then $|x+y| = x+y \leq x+y-y = x = |x| \leq |x|+|y|$. For the proof of reverse triangle inequality, we employ the triangle inequality:

$$|x| = |x+y-y| \leq |x-y|+|y|$$

and thus

$$|x|-|y| \leq |x-y|.$$

By changing the role of $x$ by $y$ in the above inequality, we obtain the other half of the inequality necessary in order to claim the reverse triangle inequality. □

**Remark 7.** It turns out that the simple-looking *triangle inequality* is extremely important and will be frequently used in the following way. Given real numbers $x, y$, and $z$, we can certainly write

$$|x-y| = |x-z+z-y|.$$

However, if we apply the triangle inequality,

$$|x-z+z-y| \leq |x-z|+|z-y|,$$

thus we obtain

$$|x-y| \leq |x-z|+|z-y|.$$

Now the expression $|x-z|$ which is equal to $|z-x|$ is best understood as the distance between $x$ and $z$. Similarly $|z-y|$ is the distance between $z$ and $y$, and $|x-y|$ is the distance between $x$ and $y$. When we think of $x$, $y$, and $z$ as points in the plane as opposed to on the real line, we see why it is referred to as the triangle inequality. In this case, the triangle inequality expresses the familiar fact that the sum of lengths of two sides of a triangle is always at least the length of the third side.

**Remark 8.** Now that we understand the fact that $|x-r|$ is the distance between $x$ and $r$, one can also state the above Theorem 9 as follows:

$$\text{If } x \in \mathbb{R} \text{ and } \epsilon > 0, \text{ there is an } r \in \mathbb{Q} \text{ such that } |x-r| < \epsilon.$$

We leave it as an exercise to the reader to check that $\mathbb{Q}$ is dense in $\mathbb{R}$ implies that every real number is a limit of rational numbers. However, to prove this, one needs the concept of a limit of a sequence which will be covered in the next section.

## Exercises

1. Compute, without proofs, the suprema and infima of the following sets:

   a) $\{n \in \mathbb{N} : \ n^2 < 20\}$.

   b) $\left\{ \dfrac{n}{n+m} : \ m, n \in \mathbb{N} \right\}$.

   c) $\{r \in \mathbb{Q} : \ r < 5\}$.

   d) $\{r \in \mathbb{Q} : \ r^2 < 5\}$.

   e) $\displaystyle\bigcup_{n=1}^{\infty} \left[ \frac{1}{n}, 2 - \frac{1}{n} \right]$.

2. Assume that $A$ and $B$ are nonempty, bounded above, and satisfy $B \subseteq A$. Show that $\sup B \leq \sup A$. What is the relationship between $\inf A$ and $\inf B$?

3. Let $A$ be a nonempty subset of $\mathbb{R}$ bounded above. Set $B = \{-a : \ a \in A\}$. Show that $B$ is bounded below and $\inf B = -\sup A$.

4. Let $x$ be a nonzero rational number and $y$ be irrational. Prove that $xy$ is irrational.

5. Show that the product of two irrational numbers may be rational or irrational.

6. Prove that for each $x \in \mathbb{R}$ and each $n \in \mathbb{N}$ there exists a rational $r_n$ such that $|x - r_n| < \frac{1}{n}$.

7. Prove that the set of irrational numbers is dense in $\mathbb{R}$.

8. Show that $\displaystyle\bigcap_{n=1}^{\infty} \left(0, \frac{1}{n}\right) = \emptyset$ (this shows that intervals in the Nested Intervals Property must be *closed* for the conclusion of the theorem to hold).

9. Suppose $A$ and $B$ are a nonempty subset of $\mathbb{R}$. Let

$$C = \{x + y : \ x \in A \text{ and } y \in B\}.$$

If $A$ and $B$ have suprema, then show that $C$ has a supremum and $\sup C = \sup A + \sup B$. What can you say about $\inf C$?

10. Suppose $A$ and $B$ are a nonempty subset of $\mathbb{R}$. Let

$$C = A \cdot B = \{ab : \ a \in A, \text{ and } b \in B\}.$$

If $A$ and $B$ have suprema, then show that $C$ has a supremum and $\sup C = \sup A \sup B$. What can you say about $\inf C$?

11. Prove that if $a$ is a real number, then there exists an integer $N$ such that $N - 1 \leq a < N$.

12. Show that:

a) The product of two irrational numbers may be irrational or rational.

b) Any irrational number multiplied by any nonzero rational number is irrational.

13. If $\alpha$ is irrational, are there any rational numbers $p/q$ such that

$$|\alpha - p/q| < 1/q^2 \ ?$$

Hint: Use the Dirichlet theorem on rational approximation of any real number.

## 1.4 Sequences and Series

Most concepts in analysis can be reduced to statements about the behavior of sequences and series. Moreover, understanding infinite series depends on understanding of sequences. A *sequence* of real numbers can be defined as a mapping from $\mathbb{N}$ into $\mathbb{R}$, a real-valued function defined on the positive integers. Although there are different notations for describing sequences, they are usually written as

$$\{x_n\} = \{x_1, x_2, \dots\},$$

where $x_n \in \mathbb{R}$ for each $n \in \mathbb{N}$.

**Example 20.** Each of the following is ways to describe a sequence.

a) $\left\{1,\ \frac{1}{2},\ \frac{1}{3},\ \frac{1}{4},\ \cdots\right\}$.

b) $x_n = (-1)^{n+1}$ or $\{x_n\} = \{1,\ -1,\ 1,\ -1,\ \dots\}$.

c) $\left\{\frac{1+n}{n}\right\}_1^{\infty} = \frac{2}{1},\ \frac{3}{2},\ \frac{4}{3},\ \dots.$

d) $\{x_n\}$, where $x_n = \sin n$ for each $n \in \mathbb{N}$.

e) $x_1 = 1; \quad x_2 = 1; \quad x_n = x_{n-2} + x_{n-1}$ if $n \geq 3$ (*Fibonacci sequence*, defined recursively).

**Definition 10** (Convergence of a Sequence)**.** The sequence $\{x_n\}$ is said to *converge to a limit* $L \in \mathbb{R}$ if for each $\epsilon > 0$ there corresponds a number $N = N(\epsilon)$ such that

$$|x_n - L| < \epsilon \quad \text{for all} \quad n \geq N.$$

In this case $L$ is the *limit*, and we write $\lim_{n\to\infty} x_n = L$ or just $x_n \to L$. If no such $L$ exists, we say $\{x_n\}$ *diverges*. The whole definition can also be written in a compact way as follows (Figure 1.15):

$$\forall \epsilon > 0, \quad \exists N \in \mathbb{N} \quad \text{such that} \quad n \geq N \Longrightarrow |x_n - L| < \epsilon.$$

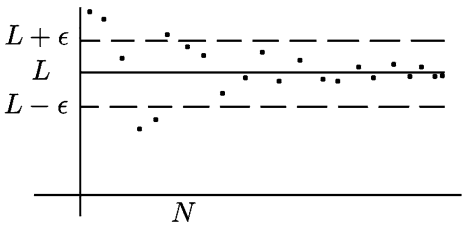



Figure 1.15: Convergence to a limit $L$

The above definition is subtle and requires an explanation. The Greek letter $\epsilon$ measures the "nearness" of the $x_n$'s to $L$, $N$ is the "stopping" place, and $|x_n - L|$ measures the distance between $x_n$ and $L$. Thus the convergence of $x_n$ to $L$ means that for each prescribed discrepancy $\epsilon$, the numbers $x_n$ are within distance $\epsilon$ of $L$ for all $n \geq N$. Thus $N$ is where you must stop for this desirable outcome to emerge. Notice that $|x_n - L| < \epsilon$ for all but a *finite* number of indices $n$. There is another issue in this definition that we must be aware of it. From some point on *every* element of the sequence approximates the limit $L$ to any desired accuracy. Therefore we could consider only values of $\epsilon$ of the form $\frac{1}{2}10^{-k}$. Then the statement $|x_n - L| < \frac{1}{2}10^{-k}$ means that $x_n$ and $L$ agree to $k$ decimal places. In other words, a sequence $\{x_n\}$ converges to $L$ precisely when eventually all the terms of the sequence agree with $L$ to $k$ decimal places.

**Example 21.** Let $x_n = \dfrac{2n+3}{n+5}$. Show that $\{x_n\}$ converges.

First we need a candidate for $L$. By plugging in numbers for $n$, it is not hard to see that when $n \to \infty$, we have $L = 2$, thus $\lim_{n\to\infty} x_n = 2$. Before we try to prove this by using the $\epsilon - N$ "game," we must explore the relationship between them. We manipulate the desired inequality in search of a suitable $N$. Now,

$$|x_n - L| = \left|\frac{2n+3}{n+5} - 2\right| = \left|\frac{2n+3-2(n+5)}{n+5}\right| = \left|\frac{-7}{n+5}\right| = \frac{7}{n+5}$$

and $\dfrac{7}{n+5} < \epsilon$ holds true if and only if $\dfrac{7}{\epsilon} < n+5$ or $\dfrac{7}{\epsilon} - 5 < n$. Thus, the number $N = \dfrac{7}{\epsilon} - 5$ does the job. Now we are ready to write an "$\epsilon - N$" proof of the convergence.

*Proof.* Let $\epsilon > 0$ be given. Setting $N = \dfrac{7}{\epsilon} - 5$ and $n \in \mathbb{N}$ with $n \geq N$, we obtain

$$|x_n - L| = \left|\frac{2n+3}{n+5} - 2\right| = \frac{7}{n+5} < \frac{7}{N+5} = \frac{7}{\dfrac{7}{\epsilon} - 5 + 5} = \epsilon. \qquad \square$$

Note that if $\epsilon = \frac{1}{10}$, then the stopping place is $N = 65$. Thus for this particular $\epsilon$, the terms of the sequence $x_n$ are within $\frac{1}{10}$ if $n \geq 65$. It should be apparent that *the value $N$ depends on the choice of $\epsilon$.*

**Definition 11.** Given a real number $L \in \mathbb{R}$ and a positive number $\epsilon > 0$, the set

$$V_\epsilon(L) = \{x \in \mathbb{R} : \ |x - L| < \epsilon\}$$

is called the *$\epsilon$-neighborhood* of $L$.

Notice that $V_\epsilon(L)$ consists of all those points whose distance from $L$ is less than $\epsilon$. Equivalently $V_\epsilon(L) = (L - \epsilon, L + \epsilon)$ is an interval since

$$|x - L| < \epsilon \Leftrightarrow L - \epsilon < x < L + \epsilon.$$

Using the above terminology we can make a topological version of the convergence of sequences. Namely, we say a sequence $\{x_n\}$ converges to $L$ if, given every $\epsilon$-neighborhood $V_\epsilon(L)$ of $L$, there exists a point in the sequence after which all the terms are in $V_\epsilon(L)$. In other words $V_\epsilon(L)$ contains all but a finite number of the terms of $\{x_n\}$. The natural number $N$ in the above definition is the last stopping place, where the sequence enters $V_\epsilon(L)$ never to leave (Figure 1.16).

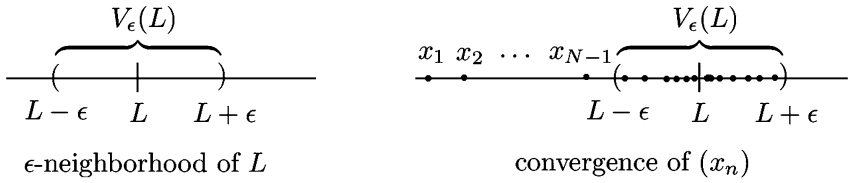



Figure 1.16: Convergence of a sequence

**Example 22.** Show that the sequence $\{x_n\} = (-1)^{n+1} = \{1, -1, 1, -1, \cdots\}$ diverges.

*Proof.* Clearly this sequence oscillates between 1 and $-1$. As in Example 21 we invoke the definition for $\epsilon$-$N$ convergence. We prove that the limit does not exist by contradiction. Suppose the sequence $\{x_n\}$ converges to $L$. From the

definition for every $\epsilon > 0$, we must find $N$. Thus let $\epsilon = \dfrac{1}{100} > 0$, and choose $N$ as in the definition. Then for any even integer $n_1$ with $n_1 \geq N$, we have

$$|x_{n_1} - L| = |-1 - L| < 0.01,$$

which means $L \in (-1.01, -0.99)$. Similarly if $n_2$ is any odd integer with $n_2 \geq N$, then we have

$$|x_{n_2} - L| = |1 - L| < 0.01$$

implying $L \in (0.99, 1.01)$. Since $L$ cannot lie in two disjoint intervals, this is a contradiction to our assumption. □

A sequence $\{x_n\}$ is *bounded above* if there exists $M$ such that $x_n \leq M$ for all $n \in \mathbb{N}$; it is *bounded below* if there exists $m$ such that $x_n \geq m$ for all $n$. The *range* of a sequence is its set of values $\{x_n : n \in \mathbb{N}\}$. We say a sequence $\{x_n\}$ is **bounded** if its range is a bounded set. In other words it is both bounded above and below. We now give two important properties of convergent sequences.

**Theorem 10.** *Let $\{x_n\}$ be a sequence of real numbers.*

*a) If $\{x_n\}$ converges to $L$ and $L^*$, then $L = L^*$ (limit is unique).*

*b) If $\{x_n\}$ is convergent, then $\{x_n\}$ is bounded.*

*Proof.*

a) Let $\epsilon > 0$ be given. There exist integers $N$ and $N'$ such that

$$n \geq N \quad \text{implies} \quad |x_n - L| < \frac{\epsilon}{2}$$

$$n \geq N' \quad \text{implies} \quad |x_n - L^*| < \frac{\epsilon}{2}.$$

Hence if $n \geq \max\{N, N'\}$, we have

$$|L - L^*| \leq |L - x_n| + |x_n - L^*| < \epsilon.$$

Since $\epsilon$ was arbitrary, we have $|L - L^*| = 0$ and therefore $L = L^*$.

b) Suppose $\{x_n\}$ converges to $L$. For $\epsilon = 1$, there is an integer $N$ such that $n \geq N$ implies that $|x_n - L| < 1$. Setting

$$K = \max\{1, |x_1 - L|, |x_2 - L|, \ldots, |x_N - L|\}$$

will give $|x_n - L| \leq K$ for $n = 1, 2, 3, \ldots$. □

There are ways to create new sequences from "known" sequences as described in the following theorem.

**Theorem 11** (Algebraic Limit Theorem)**.** *Suppose $\{x_n\}$ and $\{y_n\}$ are sequences such that $x_n \to L$ and $y_n \to M$, then*

*a)* $(x_n + y_n) \to L + M$.

*b)* $cx_n \to cL$ *for any constant* $c$.

*c)* $x_n y_n \to LM$.

*d)* $\dfrac{1}{x_n} \to \dfrac{1}{L}$ *provided* $x_n \neq 0$ *for all* $n$ *and* $L \neq 0$.

*Proof.* The proof of parts a) and b) is left as an exercise. For the proof for part c), since under suitable conditions we would like to show that $|x_n y_n - LM| < \epsilon$, we start by examining the identity

$$x_n y_n - LM = (x_n - L)(y_n - M) + L(y_n - M) + M(x_n - L).$$

Given $\epsilon > 0$, there are integers $N_1$ and $N_2$, such that

$$n \geq N_1 \quad \text{implies} \quad |x_n - L| < \sqrt{\epsilon}$$

$$n \geq N_2 \quad \text{implies} \quad |y_n - M| < \sqrt{\epsilon}.$$

If we take $N = \max\{N_1, N_2\}$, $n \geq N$ then implies

$$|(x_n - L)(y_n - M)| = |(x_n - L)||(y_n - M)| < \epsilon,$$

so that $(x_n - L)(y_n - M) \to 0$. If we apply parts a) and b) in the above expression for $x_n y_n - LM$, we can conclude that $x_n y_n \to LM$.

For part d) given $\epsilon > 0$, we must find $N$ so that

$$\left| \frac{1}{x_n} - \frac{1}{L} \right| = \left| \frac{x_n - L}{x_n L} \right| < \epsilon.$$

If we can prove that the sequence $\left| \dfrac{1}{x_n L} \right|$ is bounded, say there is a constant $M$ such that $\left| \dfrac{1}{x_n L} \right| \leq M$, then

$$\left| \frac{1}{x_n} - \frac{1}{L} \right| = \left| \frac{x_n - L}{x_n L} \right| < M|x_n - L|,$$

and we can apply part b) of the above theorem to complete the proof. To show $\left| \dfrac{1}{x_n L} \right|$ is bounded, observe that the set $\{x_n\}$ is bounded away from zero, i.e., there is $m > 0$ such that $|x_n| \geq m$ for all $n$. This is so because $x_n \to L$ implies there is an integer $N$ so that $|x_n - L| < \epsilon$ when $n \geq N$. If we let $\epsilon = \dfrac{|L|}{2} > 0$, then $|x_n| > \dfrac{|L|}{2}$ for all such $n$. The remaining set $\{|x_1|, |x_2|, \ldots, |x_N|\}$ is a finite set of positive numbers, so it has a positive minimum. Using these facts one can find a suitable value for $m$. Now it is the case that $\dfrac{1}{|x_n L|} < \dfrac{2}{m|L|}$ for all $n$. $\square$

**Remark 9.** Part d) of the above theorem implies that if $\{x_n\}$ and $\{y_n\}$ are sequences such that $x_n \to L$ and $y_n \to M$, then

$$\frac{x_n}{y_n} \to \frac{L}{M}$$

provided $y_n \neq 0$ for all $n$ and $M \neq 0$. This is the case because the result for product of sequences implies

$$x_n \frac{1}{y_n} \to L \frac{1}{M}.$$

**Theorem 12** (The Squeeze Principle)**.** *Let $\{x_n\}$, $\{y_n\}$, and $\{z_n\}$ be sequences such that*

$$x_n \leq y_n \leq z_n \qquad \text{for all } n.$$

*If $x_n \to L$ and $z_n \to L$, then $y_n \to L$ too.*

*Proof.* Since $x_n \to L$ and $z_n \to L$, for a given $\epsilon > 0$, we can choose $N_1$ such that $n \geq N_1$ implies that $|x_n - L| < \epsilon$; similarly, there is $N_2$ such that $n \geq N_2$, which implies that $|z_n - L| < \epsilon$. Now let $N = \max\{N_1, N_2\}$. Then for the sequence $y_n$, since $n \geq N$, we have that

$$L - \epsilon < x_n \leq y_n \leq z_n < L + \epsilon,$$

which implies for a given $\epsilon > 0$ there is an $N$ such that $n \geq N$ implies

$$|y_n - L| < \epsilon.$$

□

**Example 23.** Consider the sequence $\left\{\frac{\sin n}{n}\right\}$. Since $|\sin n| \leq 1$, we know

$$-\frac{1}{n} \leq \frac{\sin n}{n} \leq \frac{1}{n}.$$

Both the right- and left-hand sequences converge to zero as $n \to \infty$; the sequence in the middle is "squeezed" to the same limit.

**Remark 10.** The squeeze principle can be used to conclude more convergence properties for sequences. For example, if $x_n \to L$, then $|x_n| \to |L|$ too. Notice that $x_n \to L$ implies $x_n - L \to 0$. The *reverse triangle inequality* implies

$$0 \leq ||x_n| - |L|| \leq |x_n - L|.$$

The right-hand sequence tends to zero so that by the squeeze principle $|x_n| \to |L|$. Similarly if $|x_n| \to 0$ and $\{y_n\}$ is another *bounded* sequence, then $x_n y_n \to 0$ as well. This follows from the inequality

$$0 \leq |x_n y_n| \leq M|x_n|,$$

where the boundedness of $\{y_n\}$ implies we can choose $M$ so that $|y_n| \leq M$ for all $n$, and $|x_n y_n| \to 0$ if and only if $x_n y_n \to 0$.

We showed in Theorem 10, part b) that convergent sequences are bounded. The converse of this statement is not true. However, if a bounded sequence is monotone, then in fact it converges. Recall that a sequence is called *monotonic* if it is either nondecreasing or nonincreasing; that is,

$$x_1 \leq x_2 \leq x_3 \leq \dots \qquad \text{or} \qquad x_1 \geq x_2 \geq x_3 \geq \dots .$$

The following Monotone Convergence Theorem is very useful because it simply asserts the convergence of a sequence without explicitly finding its actual limit.

**Theorem 13** (Monotone Convergence Theorem)**.** *Every monotonic and bounded sequence of real numbers is convergent.*

*Proof.* Let $\{x_n\}$ be monotone and bounded. Suppose $\{x_n\}$ is nondecreasing (the nonincreasing case is handled similarly). Then by hypothesis its range $\{x_n : \ n \in \mathbb{N}\}$ is bounded, so we can set $\lambda = \sup\{x_n : \ n \in \mathbb{N}\}$. We need a candidate for the limit, and it is reasonable to claim $\lim\limits_{n\to\infty} x_n = \lambda$. To prove this, let $\epsilon > 0$. Since $\lambda$ is an upper bound for $\{x_n : \ n \in \mathbb{N}\}$, we have $x_n \leq \lambda$ for all $n \in \mathbb{N}$, but $\lambda - \epsilon$ is not an upper bound for each $\epsilon > 0$, i.e., there is a point in the sequence $x_N$ such that $x_N > \lambda - \epsilon$ (Figure 1.17).

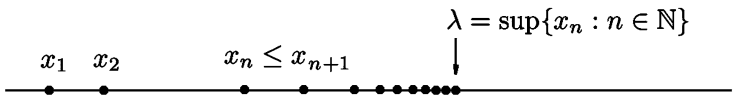



Figure 1.17: Monotone and bounded sequence

Since the sequence $\{x_n\}$ is nondecreasing, it follows that $x_n \geq \lambda - \epsilon$ for all $n \geq N$. Combining this with the inequality $x_n \leq \lambda$,

$$\lambda - \epsilon < x_N \leq x_n \leq \lambda < \lambda + \epsilon$$

implies $|x_n - \lambda| < \epsilon$ for all $n \in \mathbb{N}$, which gives $x_n \to \lambda$ as desired. □

Later we will see the Monotone Convergence Theorem will be a great help for the study of infinite series. In the coming sections we will consider certain properties of the real line that may be defined either in terms of subsets of $\mathbb{R}$ or in terms of sequences. The Nested Interval Theorem is useful in this regard. In our discussion of sequences so far we considered sequences of real numbers. However, one can have sequences in $\mathbb{R}^2$ or $\mathbb{R}^3$ or in general in the $n$-dimensional vector space $\mathbb{R}^n$ as well. In the following, we consider sequences in $\mathbb{R}^n$ and show the connection of sequences in $\mathbb{R}$ and $\mathbb{R}^n$.

**Definition 12. Euclidean $n$-space**, denoted $\mathbb{R}^n$, consists of all ordered $n$-tuples of real numbers.

$$\mathbb{R}^n = \{(x_1, \dots, x_n) : \quad x_1, \dots, x_n \in \mathbb{R}\}.$$

Clearly $\mathbb{R}^n = \mathbb{R} \times \cdots \times \mathbb{R}$ ($n$-times) is the Cartesian product of $\mathbb{R}$ with itself $n$-times. Elements of $\mathbb{R}^n$ are usually denoted by single letters that stand for $n$-tuples such as $x = (x_1, \ldots, x_n)$, and we speak of $x$ as a **point** in $\mathbb{R}^n$. Addition and scalar multiplication of $n$-tuples are defined as

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n)$$

and

$$a(x_1, \ldots, x_n) = (ax_1, \ldots, ax_n) \quad \text{for} \quad a \in \mathbb{R}.$$

The **distance** or the **norm** between two elements of $x, y \in \mathbb{R}^n$ is a real number

$$||x - y|| = \left( \sum_{i=1}^{n} (x_i - y_i)^2 \right)^{\frac{1}{2}}.$$

**Proposition 1.** *If $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ are vectors in $\mathbb{R}^n$ and we let $\rho(x, y) = \max\{|x_1 - y_1|, |x_2 - y_2|, \ldots, |x_n - y_n|\}$, then*

$$\rho(x, y) \leq ||x - y|| \leq \sqrt{n}\rho(x, y).$$

*Proof.* Since

$$|x_i - y_i| = \sqrt{|x_i - y_i|^2} \leq \sqrt{\sum_{i=1}^{n} |x_i - y_i|^2} = ||x - y||,$$

we have $\rho(x, y) \leq ||x - y||$, and

$$||x - y|| = \sqrt{\sum_{i=1}^{n} |x_i - y_i|^2} \leq \sqrt{\sum_{i=1}^{n} \max_i |x_i - y_i|^2} \leq \sqrt{n\rho(x, y)^2} = \sqrt{n}\rho(x, y)$$

gives the right-hand inequality. □

We will use the above proposition to prove the following theorem.

**Theorem 14** (Convergence in $\mathbb{R}^n$)**.** *Suppose $\{x_k\}$ is a sequence in $\mathbb{R}^n$ and $x_k = (x_k^1, x_k^2, \ldots, x_k^n)$ for $k \geq 1$. Then $\{x_k\}$ converges to $x$ in $\mathbb{R}^n$ if and only if each sequence of coordinates converges to the corresponding coordinate of $x$ as a sequence in $\mathbb{R}$. That is,*

$$x_k \to x \quad \text{in } \mathbb{R}^n \quad \text{if and only if} \quad \lim_{k \to \infty} x_k^i \to x^i \quad \text{in } \mathbb{R} \quad \text{for each } i = 1, 2, \ldots, n.$$

Note that this theorem can be written compactly as

$$\lim_{k \to \infty} (x_k^1, \ldots, x_k^n) = \left( \lim_{k \to \infty} x_k^1, \ldots, \lim_{k \to \infty} x_k^n \right).$$

*Proof.* Let $\rho(x,y) = \max\{|x_1 - y_1|, |x_2 - y_2|, \ldots, |x_n - y_n|\}$, then

$$\rho(x,y) \leq ||x - y|| \leq \sqrt{n}\rho(x,y)$$

by the above proposition. Suppose $\{x_k\}$ converges to $x$ in $\mathbb{R}^n$ and $1 \leq j \leq n$. Let $\epsilon > 0$ be given; then there is an $N$ such that $||x_k - x|| < \epsilon$ whenever $k \geq N$. Therefore, for such $k$ we have

$$|x^j - x_k^j| \leq \rho(x, x_k) \leq ||x - x_k|| < \epsilon,$$

and thus $x_k^j \to x^j$ in $\mathbb{R}$.
To prove the converse suppose that for each $j = 1, 2, \ldots, n$, we have $\lim_{k\to\infty} x_k^j \to x^j$ in $\mathbb{R}$. Let $\epsilon > 0$. For any $\epsilon_* > 0$, there exist integers $N_1, N_2, \ldots, N_n$ such that

$$\begin{aligned} |x^1 - x_k^1| &< \epsilon_* \quad \text{whenever} \quad k \geq N_1 \\ |x^2 - x_k^2| &< \epsilon_* \quad \text{whenever} \quad k \geq N_2 \\ &\vdots \qquad\qquad\quad \vdots \\ |x^n - x_k^n| &< \epsilon_* \quad \text{whenever} \quad k \geq N_n. \end{aligned}$$

Now set $N = \max\{N_1, N_2, \ldots, N_n\}$. Then for $k \geq N$,

$$||x - x_k|| \leq \sqrt{n}\rho(x, x_k) \leq \sqrt{n}\epsilon_*.$$

Thus by taking $\epsilon_* = \dfrac{\epsilon}{\sqrt{n}}$, we obtain

$$||x - x_k|| < \epsilon \quad \text{whenever } k \geq N$$

proving $x_k \to x \quad$ in $\mathbb{R}^n$. □

**Example 24.** Consider the sequence $\{x_k\}$ in $\mathbb{R}^2$, where $x_k = \left(\dfrac{1}{k^2}, \dfrac{1}{k^3}\right)$. The components of this sequence are $\dfrac{1}{k^2}$ and $\dfrac{1}{k^3}$, and each converges to zero; then, by the above theorem the sequence $x_k = \left(\dfrac{1}{k^2}, \dfrac{1}{k^3}\right) \to (0,0)$ in $\mathbb{R}^2$.

**Remark 11.** To describe the convergence of a sequence $\{x_n\}$ to $x_0$ in $\mathbb{R}^2$, one can consider the open $\epsilon$-neighborhood of the point $x_0$. Recall that $U_\epsilon(x_0)$, an open neighborhood of $x_0$ in $\mathbb{R}^2$, is defined as

$$U_\epsilon(x_0) = \{x \in \mathbb{R}^2 : ||x - x_0|| < \epsilon\}.$$

By $||\cdot||$ we mean the Euclidean distance in $\mathbb{R}^2$. As one can see from the following Figure 1.18, $U_\epsilon(x_0)$ contains all but a finite number of terms of $\{x_n\}$.

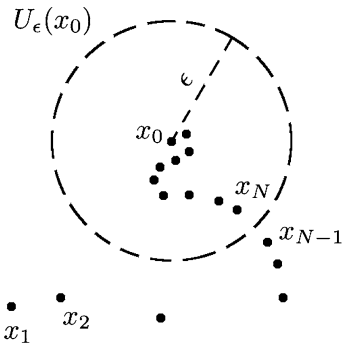



Figure 1.18: Convergence of a sequence in $\mathbb{R}^2$

Much of our discussion of sequences in $\mathbb{R}$ still makes sense in $\mathbb{R}^n$, but some do not. For example, there is no natural order imposed on $\mathbb{R}^n$, and so the discussion of monotone sequences does not apply in this context.

## Subsequences

**Definition 13.** Given a sequence $\{x_n\}$ of real numbers, consider a sequence $\{n_k\}$ of positive integers such that $n_1 < n_2 < n_3 < \cdots$; then the sequence

$$x_{n_1}, x_{n_2}, x_{n_3}, \ldots$$

is called a **subsequence** of $\{x_n\}$ and is denoted by $\{x_{n_k}\}$, where $k \in \mathbb{N}$ indexes the subsequence.

Any given sequence $\{x_n\}$ has many subsequences, and note that the order of the terms in a subsequence is the same as in the original sequence and repetitions are not allowed. Subsequences may or may not behave like the original sequence. For example, if $\{x_n\} = 1, -1, 1, -1, \ldots$, then $\{x_n\}$ has no single limit; thus it is divergent, but the subsequences

$$x_1, x_3, x_5, \ldots = 1, 1, 1, \ldots \qquad x_2, x_4, x_6, \cdots = -1, -1, -1, \ldots$$

converge to 1 and $-1$, respectively. On the other hand, the sequence $\{x_n\} = \left\{\dfrac{1}{10^n}\right\}$ converges to zero and so does every one of its subsequences. These examples lead us to the following:

**Proposition 2.** *Let $\{x_n\}$ be a sequence and $L$ be a real number.*

*a) If $\{x_n\}$ converges to $L$, then every subsequence $\{x_{n_k}\}$ converges to $L$ too.*

*b) If $\{x_n\}$ has subsequences converging to different limits, then $\{x_n\}$ diverges.*

*Proof.* To prove a) let $\epsilon > 0$ be given; since $\{x_n\}$ converges to $L$, there exists $N$ such that $|x_n - L| < \epsilon$ whenever $n \geq N$. Note that for any subsequence, $n_k \geq k$, so this very same $N$ also works for the subsequence $\{x_{n_k}\}$. If $k > N$, then

$$n_k \geq k > N, \quad \text{and} \quad |x_{n_k} - L| < \epsilon.$$

The proof of part b) follows from part a). Indeed, $\{x_n\} \to L$ if and only if all subsequences of $\{x_n\}$ converge to $L$. □

**Theorem 15** (The Bolzano–Weierstrass Theorem)**.** *Every bounded sequence of real numbers has a convergent subsequence.*

*Proof.* Suppose $\{x_n\}$ is a bounded sequence in $\mathbb{R}$ so that there exists $M > 0$ with $-M \leq x_n \leq M$ for every $n$. Bisect the interval $[-M, M]$ into two closed intervals $[-M, 0]$ and $[0, M]$. At least one of these closed intervals must contain infinitely many $x_n$. Select a half for which this is the case label that interval by $I_0$, and select $n_0$ for which $x_{n_0} \in I_0$. Next, split $I_0$ into closed intervals of equal length, and let $I_1$ be the half that again contains infinitely many points of $x_n$. As there are infinitely many $x_n$'s available, we can select $n_1 > n_0$ with the property that $x_{n_1} \in I_1$ (Figure 1.19).

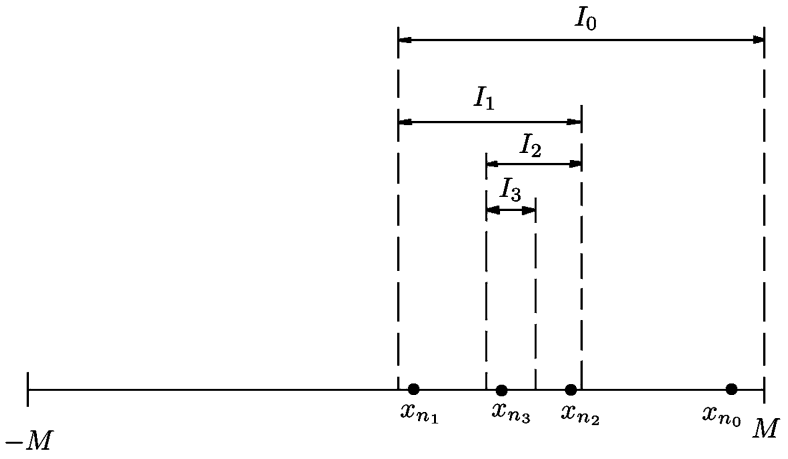



Figure 1.19: The bisection process used in the proof of Bolzano–Weierstrass Theorem

Continue these processes to obtain subintervals, indices, and points. From this bisection processes we obtain the following:

a) $I_0 \supseteq I_1 \supseteq I_2 \supseteq \cdots$ (a nested set of closed and bounded intervals).

b) The length of $I_k$ is $\frac{M}{2^k}$.

c) Increasing indices $n_0 < n_1 < n_2 < \cdots$ and a subsequence $\{x_{n_k}\}$ such that $x_{n_k} \in I_k$.

Next, we claim that this subsequence $\{x_{n_k}\}$ is the convergent subsequence we have been looking for. However, we must first catch its limit $x$. The Nested Intervals Property comes in handy here, which guarantees the existence of at least one point $x \in \mathbb{R}$ contained in every $I_k$, i.e., $\bigcap_{n=1}^{\infty} I_n \neq \emptyset$. We claim $x_{n_k} \to x$. If we label $I_k = [a_k, b_k]$ and consider the left hand of these intervals, we see a sequence $a_1, a_2, a_3, \ldots$ Since $I_{k+1} \subseteq I_k \subseteq [-M, M]$, we have

$$-M \leq a_k \leq a_{k+1} \leq M \quad \text{for every } k.$$

This sequence is monotone increasing and bounded, so by the Monotone Convergence Theorem it must converge to some number $x$. Given $\epsilon > 0$, there is an integer $K_1$ such that $|a_k - x| < \epsilon$ whenever $k \geq K_1$. Now observe that for each $k$, we have

$$|x_{n_k} - x| \leq |x_{n_k} - a_k| + |a_k - x| \leq \frac{M}{2^k} + |a_k - x|.$$

Let $\epsilon > 0$. By construction the length of $I_k = b_k - a_k = \dfrac{M}{2^k}$ and $\dfrac{1}{2^k} \to 0$ as $k \to \infty$. Therefore, there is an integer $K_2$ such that $\dfrac{M}{2^k} < \dfrac{\epsilon}{(2M)}$ whenever $k \geq K_2$. Let $K = \max\{K_1, K_2\}$; thus if $k \geq K$, we must have

$$|x_{n_k} - x| \leq \frac{M}{2^k} + |a_k - x| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \qquad \square$$

## Cauchy Sequences

**Definition 14.** A **Cauchy sequence** of real numbers is a sequence $\{x_n\}$ with the property that for each $\epsilon > 0$ there is an integer $N$ depending on $\epsilon$ such that

$$|x_n - x_m| < \epsilon \quad \text{for all} \quad n, m \geq N.$$

Notice that this definition is similar to the definition of a convergent sequence, but there is no mention of the limit $L$. Furthermore, the definition of a Cauchy sequence suggests that the terms of the sequence are "bunching up," but it requires that **all** terms with large enough index are close to one another, not just consecutive terms. Sometimes in the definition the condition $n > m \geq N$ is used. This does not change anything since $n = m$ is the trivial case. Consider the sequence $\left\{\dfrac{1}{n}\right\}$; clearly it is convergent, and we can guess that terms of this sequence "bunch up." Let $\epsilon > 0$ be given, set $N = \dfrac{1}{\epsilon}$, and since if $n > m > N$, then

$$|x_n - x_m| = \left|\frac{1}{m} - \frac{1}{n}\right| = \frac{1}{m} - \frac{1}{n} < \frac{1}{m} < \frac{1}{\frac{1}{\epsilon}} = \epsilon.$$

There is nothing special about the convergent sequence $\left\{\dfrac{1}{n}\right\}$, since, as shown in the next lemma, convergent sequences are Cauchy sequences.

**Lemma 3.** *Every convergent sequence is a Cauchy sequence.*

*Proof.* Assume $\{x_n\}$ converges to $L$. Then given $\epsilon > 0$, we can choose $N$ such that $|x_n - L| < \dfrac{\epsilon}{2}$ whenever $n \geq N$. Now we apply the triangle inequality to $|x_n - x_m|$ and observe that for $n > m > N$, we have

$$|x_n - x_m| \leq |x_n - L| + |x_m - L| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Thus $\{x_n\}$ is a Cauchy sequence. □

As expected, if both $x_n$ and $x_m$ are close to a limit $L$, then they are close to each other; thus every convergent sequence is Cauchy. The converse, however, is not so obvious. In fact for a Cauchy sequence to be convergent, we must have a limit $L$ to approach. We will use the Bolzano–Weierstrass Theorem for the converse. For this purpose we first claim that Cauchy sequences are bounded.

**Lemma 4.** *Every Cauchy sequence is bounded.*

*Proof.* Choose $\epsilon = 1$, and refer to the definition of the Cauchy sequence to conclude that for some integer $N$ the inequality $|x_m - x_n| < 1$ holds for all $n, m \geq N$. Thus we must have

$$|x_n| \leq |x_N| + |x_n - x_N| \leq |x_N| + 1$$

for all $n \geq N$. But this implies that the sequence is bounded, since

$$|x_n| \leq \max\{|x_1|, |x_2|, \ldots, |x_N|, |x_N| + 1\}$$

for all $n \in \mathbb{N}$. □

**Theorem 16** (Cauchy Criterion)**.** *Every Cauchy sequence of real numbers is convergent.*

*Proof.* Suppose $\{x_n\}$ is a Cauchy sequence. Lemma 4 guarantees that $\{x_n\}$ is bounded. Thus we may use the Bolzano–Weierstrass Theorem to produce a convergent subsequence $\{x_{n_k}\}$. Let $x_{n_k} \to L$. The final step is to show that if a Cauchy sequence has a convergent subsequence, then in fact the full sequence $\{x_n\}$ is convergent. Indeed, if $x_{n_k} \to L$ as $k \to \infty$, then for each $\epsilon > 0$ we have

$$|x_n - L| \leq |x_n - x_{n_k}| + |x_{n_k} - L| < |x_n - x_{n_k}| + \frac{\epsilon}{2}$$

for all sufficiently large indices $n_k$. But because $\{x_n\}$ is a Cauchy sequence,

$$|x_n - L| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

for all sufficiently large $n$. This shows that $x_n \to L$ as $n \to \infty$. □

**Remark 12** (Completeness of $\mathbb{R}$ Again)**.** In the previous section we discussed the Axioms of Completeness, which simply asserts that "every nonempty set bounded above has a supremum." We then used this axiom to prove the Nested Intervals property and the Monotone Convergence Theorem. In the proof of Bolzano–Weierstrass, we utilized both the Monotone Convergence and Nested Intervals Theorems. Finally, the Bolzano–Weierstrass Theorem implies the Cauchy Criterion for convergent sequences. All of these fundamental results come from the same family, each describing the completeness of $\mathbb{R}$ in a particular way.

## Lim Sup and Lim Inf

Note that many things we want to know about functions $f : \mathbb{R} \to \mathbb{R}$ or the real line $\mathbb{R}$ itself can be described in terms of sequences. For example, we will see that a continuous function $f : \mathbb{R} \to \mathbb{R}$ can be defined as a function that "preserves" convergent sequences: $f(\lim_{n\to\infty} x_n) = \lim_{n\to\infty} f(x_n)$. So far we have seen that monotone and bounded sequences converge, and a convergent sequence is necessarily bounded. Of course not every bounded sequence converges, but we can wonder how "far" is a bounded sequence from being convergent. Moreover, a sequence may or may not have limit, or it may have several converging subsequences. The following notions of limit superior and limit inferior will help to clarify these issues. We start with a convention for unbounded sequences; if $\{x_n\}$ is not bounded above, we write

$$\limsup_{n\to\infty} x_n = +\infty.$$

Similarly, if $\{x_n\}$ is not bounded below, we write

$$\liminf_{n\to\infty} x_n = -\infty.$$

If $x_n \to +\infty$, we say $\liminf_{n\to\infty} x_n = +\infty$, and if $x_n \to -\infty$, we write $\limsup_{n\to\infty} x_n = -\infty$.

**Definition 15.** Let $\{x_n\}$ be a sequence of real numbers. We define its limit superior by

$$\limsup_{n\to\infty} x_n := \lim_{n\to\infty} (\sup\{x_k : \ k \geq n\})$$

and its limit inferior by

$$\liminf_{n\to\infty} x_n := \lim_{n\to\infty} (\inf\{x_k : \ k \geq n\}).$$

Note that for a bounded sequence $\{x_n\}$, we can define a new sequence $\{a_n\}$ by

$$\begin{aligned} a_1 &= \sup\{x_1, x_2, x_3, \dots\}, \ a_2 = \sup\{x_2, x_3, x_4, \dots\}, \dots, \\ a_n &= \sup\{x_n, x_{n+1}, x_{n+2}, \dots\} \end{aligned}$$

and call this sequence the *sup-sequence* associated with $\{x_n\}$. Then the above definition simply says $\limsup_{n\to\infty} x_n = \lim_{n\to\infty} a_n$. Moreover, if $\{x_n\}$ is bounded above and does not tend to $-\infty$ and if we let $A$ denote the set of subsequential limits of $\{x_n\}$, the Bolzano–Weierstrass Theorem implies that $A \neq \emptyset$. We then define $\limsup_{n\to\infty} x_n = \sup A$. Similarly, if $\{x_n\}$ is bounded below and does not tend to $+\infty$, one can set $\liminf_{n\to\infty} = \inf A$. It is worth mentioning that the number $\sup A$ actually belongs to $A$; thus $\limsup_{n\to\infty} x_n$ is actually the largest subsequential limit. Similarly, $\liminf_{n\to\infty} x_n$ is the smallest subsequential limit if $\{x_n\}$ is bounded below and does not tend to $+\infty$. There are two simple consequences of the above definition. The first one is that

$$\limsup_{n\to\infty}(-x_n) = -\liminf_{n\to\infty} x_n,$$

i.e., upper and lower limits are kind of "reflections" of each other; thus properties of one will transfer to the other. The second consequence of the definition is the simple fact that

$$\liminf_{n\to\infty} x_n \leq \limsup_{n\to\infty} x_n.$$

**Remark 13.** An alternate notation is $\limsup_{n\to\infty} = \overline{\lim}$ and $\liminf_{n\to\infty} = \underline{\lim}$.

**Theorem 17.** *A sequence of real numbers $\{x_n\}$ converges to a limit $x \in \mathbb{R}$ if and only if*

$$\liminf_{n\to\infty} x_n = \limsup_{n\to\infty} x_n = x.$$

*Proof.* In case that $\{x_n\}$ converges to $x$, we know that the sequence $\{x_n\}$ is bounded and $A = \{x\}$. From the explanation above,

$$\liminf_{n\to\infty} x_n = \inf(A) = x = \sup(A) = \limsup_{n\to\infty} x_n.$$

Now suppose

$$\liminf_{n\to\infty} x_n = \limsup_{n\to\infty} x_n = x.$$

Let $A_n := \sup\{x_k : k \geq n\}$ and $B_n := \inf\{x_k : k \geq n\}$. Note that our assumption is $\lim_{n\to\infty} A_n = \lim_{n\to\infty} B_n$. Thus given $\epsilon > 0$, there is $N \in \mathbb{N}$ such that $|x - A_N| < \epsilon$ and $|x - B_N| < \epsilon$. Since $B_N \leq x_n \leq A_N$ for all $n \geq N$, we also have $-(x - B_N) \leq x_n - x \leq A_N - x$ or $|x_n - x| < \epsilon$ for all $n \geq N$. □

**Example 25.** Let $\{x_n\}$ be a sequence defined as $\{x_n\} = (-1)^n(1 + \frac{1}{n^2})$. To find its $\liminf_{n\to\infty} x_n$ and $\limsup_{n\to\infty} x_n$, notice that

$$\inf\{x_k : k \geq n\} = \begin{cases} -\left(1 + \frac{1}{n^2}\right) & \text{if n is odd} \\ -\left(1 + \frac{1}{(n+1)^2}\right) & \text{if n is even,} \end{cases}$$

hence $\liminf_{n\to\infty} x_n = -1$. Similarly,

$$\sup\{x_k : k \geq n\} = \begin{cases} 1 + \frac{1}{n^2} & \text{if n is even} \\ 1 + \frac{1}{(n+1)^2} & \text{if n is odd,} \end{cases}$$

and hence $\limsup_{n\to\infty} x_n = 1$.

## Some Standard Sequences

The following are some special sequences whose limits can be found with the help of the binomial theorem and the "squeeze principle." (See [42], p. 57.)

**Example 26.**

a) For each real number $x \in \mathbb{R}$ with $|x| < 1$, $\lim_{n\to\infty} x^n = 0$.

b) If $a > 0$, then $\lim_{n\to\infty} \sqrt[n]{a} = 1$.

c) $\lim_{n\to\infty} \sqrt[n]{n} = 1$.

d) For each pair of numbers $a$ and $x$ with $|x| < 1$, $\lim_{n\to\infty} n^a x^n = 0$.

e) For each number $a > 0$, $\lim_{n\to\infty} \dfrac{\log n}{n^a} = 0$.

## Convergence in $\mathbb{C}$

A *complex sequence* $\{z_n = x_n + iy_n\}$ is an assignment of a complex number $z_n$ to each natural number $n = 1, 2, \ldots$, where Re $z_n = x_n$ and Im $z_n = y_n$, (for a detailed study of a complex number, we refer the reader to Chapter 3). Let us examine an example.

**Example 27.** $\{i^n\}$, $\left\{\left(\dfrac{1-i}{1+i}\right)^n\right\}$, and $\{(1+i)^n + (1-i)^n\}$ are all examples of complex sequences.

Note that for any given complex sequence $\{z_n\}$, we can form two real sequences, namely $\{\text{Re } z_n\}$ and $\{\text{Im } z_n\}$, by considering just the real and imaginary parts of the sequence $\{z_n\}$. For example, if $\{z_n\} = \left\{\dfrac{1}{n} + i\left(\dfrac{n}{n+2}\right)\right\}$, then $\{\text{Re } z_n\} = \left\{\dfrac{1}{n}\right\}$ and $\{\text{Im } z_n\} = \left\{\dfrac{n}{n+2}\right\}$ are both real sequences.

**Definition 16.** The sequence $\{z_n\}$ *converges* to limit $z$ (in symbols: $z_n \to z$) as $n \to \infty$ if, given $\epsilon > 0$, there is a natural number $N = N(\epsilon)$ such that

$$|z_n - z| < \epsilon \quad \text{whenever} \quad n \geq N.$$

Here $|z_n| = \sqrt{x_n^2 + y_n^2}$ denotes the modulus of $z$.

This definition of convergence mimics those for real sequences, so the question is can one carry over most of everything to complex sequences from real sequences? Not everything! We need to be careful about the following:

**Lemma 5.** *Complex numbers cannot be ordered.*

*Proof.* Suppose such an ordering exists. Then either $i \geq 0$ or $i \leq 0$. Suppose $i \geq 0$, then $i \cdot i \geq 0$, so $-1 \geq 0$ which is absurd. Now suppose $-i \geq 0$, then $(-i) \cdot (-i) \geq 0$ or $i^2 \geq 0$ which is absurd again. Therefore, if one requires the ordering properties for reals to hold, then such ordering is impossible for $\mathbb{C}$. □

Thus proofs that depend on the order structure of $\mathbb{R}$ do not transfer directly to $\mathbb{C}$. The following proposition gives the link between convergence in $\mathbb{R}$ and convergence in $\mathbb{C}$.

**Proposition 3.** *Let $\{z_n\}$ be a sequence in $\mathbb{C}$. Then the sequence $\{z_n\}$ converges if and only if the real sequences $\{Re\ z_n\}$ and $\{Im\ z_n\}$ both converge.*

*Proof.* Let $z_n = x_n + iy_n$ and $z = x + iy$; to complete the proof all we need is the following basic inequalities:

$$|\text{Re } z_n| = |x_n| \leq |z_n| = \sqrt{x_n^2 + y_n^2} \quad \text{and} \quad |\text{Im} z_n| = |y_n| \leq |z_n| = \sqrt{x_n^2 + y_n^2}$$

and

$$|z_n - z|^2 = |(x_n + iy_n) - (x + iy)|^2 = |(x_n - x) + i(y_n - y)|^2 = (x_n - x)^2 + (y_n - y)^2.$$

□

Going back to the previous example,

$$z_n = \frac{1}{n} + i\left(\frac{n}{n+2}\right) \to 0 + i,$$

because $\dfrac{1}{n} \to 0$ and $\dfrac{n}{n+2} \to 1$.

The sequence $\{w_n\}$ is called a *subsequence* of the sequence $\{z_n\}$ if there exist natural numbers $n_1 < n_2 < \cdots$ such that $w_k = z_{n_k}$ for $k = 1, 2, \ldots$. Now we use Proposition 3 to derive a theorem from its real counterpart (see Theorem 15). We say that the sequence $\{z_n\}$ is *bounded* if there is a finite constant $M$ such that $|z_n| \leq M$.

**Theorem 18.** *Any bounded sequence in $\mathbb{C}$ has a convergent subsequence.*

*Proof.* Let $\{z_n\}$ be a bounded sequence in $\mathbb{C}$, meaning there exists $M \geq 0$ such that $|z_n| \leq M$ for all $n$. Then from the fact $|\text{Re } z_n| \leq |z_n|$, we have $|\text{Re } z_n| \leq M$, so $\{\text{Re } z_n\}$ is a bounded sequence in $\mathbb{R}$. Hence, there exist $n_1 < n_2 < \cdots$ such that the subsequence $\{\text{Re}\, z_{n_k}\}$ converges. Now consider the subsequence $\{\text{Im}\, z_{n_k}\}$, which is also a bounded real sequence and hence converges, i.e., we can choose natural numbers $m_1 < m_2 < \cdots$ with $m_j = n_{k_j}$ so that $\{\text{Im } z_{m_j}\}$

converges. Now as a subsequence of $\{\operatorname{Re} z_{n_k}\}$, the sequence $\{\operatorname{Re} z_{m_j}\}$ must converge too. Note that we are taking a subsequence of a subsequence. Now apply Proposition 3 to combine real and imaginary parts of this sequence to conclude that $\{z_{m_j}\}$ provides a convergent subsequence of $\{z_n\}$. □

**Corollary 3.** *An infinite closed and bounded subset $K$ of $\mathbb{C}$ has a limit point in $K$.*

*Proof.* Select a sequence of distinct points $z_n$ in $K$. Since $K$ is bounded, the above theorem asserts that $\{z_n\}$ has a subsequence which converges. If $z_n \to z$, then $z$ is a limit point of $K$; because $K$ is closed it contains $z$. Note that we will see later what we mean by "closed" sets in detail. Closed and bounded sets have a special name in $\mathbb{C}$: compact sets. □

## Infinite Series

An infinite series is a sum with infinitely many terms:

$$\sum_{k=1}^{\infty} a_k = a_1 + a_2 + a_3 + \cdots,$$

where the elements $a_n$ come from a number system in which addition is defined. We will fix our number system to be the set of complex numbers $\mathbb{C}$, but we give ourselves the freedom of restricting to the real numbers or even the rational numbers.

An infinite series $\displaystyle\sum_{k=1}^{\infty} a_k$ is said to *converge* to a *sum* $s$ if its sequence of *partial sums* $s_n = \displaystyle\sum_{k=1}^{n} a_k$ converges to $s$. We write

$$\sum_{k=1}^{\infty} a_k = s.$$

If $\{s_n\}$ diverges, the series is said to diverge. Note that every sequence $\{a_n\}$ has a representation as the partial sum of a series, putting $s_1 = a_1$ and $a_n = s_n - s_{n-1}$ and $s_n = \displaystyle\sum_{k=1}^{n} a_k$, and vice versa. Therefore, in a certain sense sequences and series are in one-to-one correspondence.

**Example 28.** The infinite series $\displaystyle\sum_{k=1}^{\infty} \frac{1}{k(k+1)}$ converges to 1. We see this by finding its partial sums:

$$s_n = \sum_{k=1}^{n} \frac{1}{k(k+1)} = \sum_{k=1}^{n} \left( \frac{1}{k} - \frac{1}{k+1} \right) = 1 - \frac{1}{n+1} = \frac{n}{n+1}.$$

Thus, $\displaystyle\sum_{k=1}^{\infty} \frac{1}{k(k+1)} = \lim_{n\to\infty} s_n = \lim_{n\to\infty} \frac{n}{n+1} = 1.$

However, most series do not have partial sum sequences with a nice simple formula like the one in the above example. There are two questions we need to answer about series:

a) How do we tell if a given series converges or diverges?

b) If a series does converge, how do we find its sum?

Although it requires certain work to decide if a given series converges or diverges, the second question presents much more difficulties than the first one. We now start by applying some well-known theorems about sequences to the sequence of partial sums, to obtain some useful criteria.

**Theorem 19.** *A series with nonnegative terms converges if and only if its sequence of partial sums forms a bounded sequence.*

This follows from the fact that for a series with nonnegative terms $a_k \geq 0$ for all $k$, $s_n \leq s_{n+1}$. If $\{s_n\}$ is monotonic, then $\{s_n\}$ converges if and only if it is bounded.

**Theorem 20.** *A series* $\displaystyle\sum_{k=1}^{\infty} a_k$ *converges if and only if for every* $\epsilon > 0$ *there is an integer* $N$ *such that* $\left|\displaystyle\sum_{k=n}^{m} a_k\right| \leq \epsilon$ *for* $m \geq n \geq N$.

This is a direct consequence of the Cauchy criterion for sequences. In particular, if $m = n$, then the conclusion of the above theorem becomes $|a_n| \leq \epsilon$ whenever $n \geq N$, which means that if $\displaystyle\sum_{n=1}^{\infty} a_n$ converges, then $\lim_{n\to\infty} a_n = 0$. Thus we can conclude if $\lim_{n\to\infty} a_n \neq 0$, then $\displaystyle\sum_{n=1}^{\infty} a_n$ diverges. This is called the *divergence test*. However the condition $\lim_{n\to\infty} a_n = 0$ does not ensure the convergence of $\displaystyle\sum_{n=1}^{\infty} a_n$. A typical counterexample is the series $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n}$ which we will see diverges (see Example 30 below). Another very important example of a series is the following *geometric series*:

**Example 29.** If $x \neq 1$, the geometric series $\displaystyle\sum_{k=0}^{\infty} x^k$ has partial sums

$$s_n = \sum_{k=0}^{n} x^k = \frac{1 - x^{n+1}}{1 - x}.$$

This can be proved using mathematical induction. The geometric series converges if and only if $|x| < 1$, in which case the sum is $s = \dfrac{1}{1-x}$. This is because $\lim_{n\to\infty} x^n = 0$ for $|x| < 1$ and thus

$$\lim_{n\to\infty} s_n = \lim_{n\to\infty} \frac{1-x^{n+1}}{1-x} = \frac{1}{1-x}.$$

For $x = 1$, we get

$$\sum_{k=0}^{\infty} x^k = 1 + 1 + 1 + \cdots,$$

which clearly diverges.

**Example 30** (Harmonic Series)**.** The series $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n}$ is called the *harmonic series.* Clearly its sequence of partial sums

$$s_m = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{m}$$

is increasing, but it is increasing at a slow pace, so we might think naively that this sequence of partial sums to be somewhat bounded. However,

$$s_4 = 1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) > 1 + \frac{1}{2} + \left(\frac{1}{4} + \frac{1}{4}\right) = 2,$$

and a similar calculation yields $s_8 > 1 + 3\left(\frac{1}{2}\right)$. Indeed, $s_{2^k}$ is unbounded since

$$\begin{aligned} s_{2^k} &= 1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \cdots + \frac{1}{8}\right) + \cdots + \left(\frac{1}{2^{k-1}+1} + \cdots + \frac{1}{2^k}\right) \\ &> 1 + \frac{1}{2} + \left(\frac{1}{4} + \frac{1}{4}\right) + \left(\frac{1}{8} + \cdots + \frac{1}{8}\right) + \cdots + \left(\frac{1}{2^k} + \cdots + \frac{1}{2^k}\right) \\ &= 1 + \frac{1}{2} + 2\left(\frac{1}{4}\right) + 4\left(\frac{1}{8}\right) + \cdots + 2^{k-1}\left(\frac{1}{2^k}\right) = 1 + k\left(\frac{1}{2}\right). \end{aligned}$$

Thus the sequence of partial sums associated with the harmonic series is not bounded. Because convergent sequences are bounded, the harmonic series diverges.

The following theorem called the *Cauchy Condensation Test* is remarkable in the sense that it gives convergence or divergence of $\displaystyle\sum_{n=0}^{\infty} a_n$ by looking at a subsequence $\{a_{2^k}\}$ of $\{a_k\}$.

**Theorem 21.** *Let $\{a_n\}$ be a sequence with $a_n \geq 0$ for all $n \in \mathbb{N}$ and $a_1 \geq a_2 \geq a_3 \geq \cdots$. Then the series $\displaystyle\sum_{n=0}^{\infty} a_n$ converges if and only if $\displaystyle\sum_{n=0}^{\infty} 2^n a_{2^n}$ converges.*

*Proof.* Since the series has nonnegative terms, it converges if and only if its partial sums form a bounded sequence. Thus it suffices to show boundedness of the partial sums. First assume that $\sum_{n=0}^{\infty} 2^n a_{2^n}$ converges. Then the partial sums

$$t_k = a_1 + 2a_2 + \cdots + 2^k a_{2^k}$$

are bounded; that is, there exists a positive $M > 0$ such that $t_k \leq M$ for all $k \in \mathbb{N}$. Now we need to show that the partial sums

$$s_m = a_1 + a_2 + \cdots + a_m$$

are bounded. For $m < 2^k$,

$$s_m = a_1 + (a_2 + a_3) + (a_4 + a_5 + a_6 + a_7) + \cdots + (a_{2^k} + \cdots + a_{2^{k+1}-1}).$$

Therefore,

$$s_m \leq a_1 + (a_2 + a_2) + (a_4 + a_4 + a_4 + a_4) + \cdots + (a_{2^k} + \cdots + a_{2^k}) = t_k \leq M.$$

For the proof of reverse implication, consider $m > 2^k$, and show that $2s_m \geq t_k$ to conclude that the sequences $\{s_m\}$ and $\{t_k\}$ are either both unbounded or both bounded. □

**Corollary 4.** *The series* $\sum_{n=1}^{\infty} \frac{1}{n^p}$ *converges if* $p > 1$ *and diverges if* $p \leq 1$.

*Proof.* From the well-known fact that if $\sum_{n=1}^{\infty} a_n$ converges, then $\lim_{n\to\infty} a_n = 0$, the given series diverges when $p \leq 0$. Now, if $p > 0$, then we can use the Cauchy condensation test to examine the series $\sum_{k=0}^{\infty} 2^k \frac{1}{2^{kp}} = \sum_{k=0}^{\infty} 2^{(1-p)k}$. Comparing with the geometric series by taking $x = 2^{1-p}$, we conclude that when $2^{1-p} < 1$, i.e., when $1 - p < 0$, the series converges. □

**Definition 17.** A series $\sum_{k=0}^{\infty} a_k$ is said to be *absolutely convergent* if $\sum_{k=0}^{\infty} |a_k|$ is convergent.

**Theorem 22.** *Every absolutely convergent series converges.*

*Proof.* The proof follows from the Cauchy criterion for convergence and the inequality

$$|s_m - s_n| \leq \left| \sum_{k=n}^{m} a_k \right| \leq \sum_{k=n}^{m} |a_k| = t_m - t_n.$$ □

Note that the above theorem is essentially equivalent to saying if $|a_k| \leq b_k$ and if $\sum_{k=0}^{\infty} b_k$ is convergent, $\sum_{k=0}^{\infty} a_k$ is absolutely convergent.

**Remark 14.** If $\sum_{k=0}^{\infty} a_k$ converges, but $\sum_{k=0}^{\infty} |a_k|$ diverges, we say $\sum_{k=0}^{\infty} a_k$ is *conditionally convergent* or $\sum_{k=0}^{\infty} a_k$ converges *non-absolutely.* A well-known example of a conditionally convergent series is the *alternating harmonic series* $\sum_{k=0}^{\infty} (-1)^k \frac{1}{k}$.

In the following we list the comparison, ratio, and root tests. They are tests for absolute convergence, and they cannot give information about conditionally convergent series. The basic theorem on series with alternating signs is due to G. Leibniz. For Leibniz's alternating series theorem, we refer the reader to [18], pp. 72–73.

**Theorem 23** (Comparison Test)**.** *Suppose $a_n > 0$ for all $n \in \mathbb{N}$ and $\sum_{n=0}^{\infty} a_n$ converges. If $b_n \in \mathbb{C}$ is such that*

$$|b_n| \leq a_n \qquad \text{for all } n,$$

*then the series $\sum_{n=0}^{\infty} b_n$ converges absolutely and hence converges.*

The proof is left as an exercise. Note that although the comparison test is very useful, it requires knowledge of some convergent series $\sum_{n=0}^{\infty} a_n$ ahead of time. One of the most useful series for comparison is the geometric series $\sum_{n=0}^{\infty} ar^n$. In the case $|r| < 1$, this series converges and the sum is $\sum_{k=0}^{\infty} ar^n = \frac{a}{1-r}$.

**Theorem 24** (Ratio Test)**.** *Suppose $\sum_{n=0}^{\infty} a_n$ is a series of nonzero complex numbers and suppose the limit*

$$r = \lim_{n \to \infty} \left| \frac{a_{n+1}}{a_n} \right|$$

*exists, then*

*a) If $r < 1$, $\sum_{n=0}^{\infty} a_n$ converges absolutely.*

*b) If* $r > 1$, $\sum_{n=0}^{\infty} a_n$ *diverges.*

*c) If* $r = 1$, *the test is inconclusive.*

*Proof.* The result is clear when $r > 1$. For the case when $r = \lim_{n\to\infty} \left| \frac{a_{n+1}}{a_n} \right| < 1$, if $\lambda$ satisfies $r < \lambda < 1$, then there exists $N \in \mathbb{N}$ such that

$$\frac{|a_{n+1}|}{|a_n|} < \lambda \quad \text{for all} \quad n \geq N.$$

Therefore,

$$|a_n| \leq |a_N|\lambda^{n-N} \quad \text{for all} \quad n \geq N.$$

Now apply the comparison test. The test gives no information when $r = 1$. For example, the series $\sum_{n=0}^{\infty} \frac{1}{n}$ and $\sum_{n=0}^{\infty} \frac{1}{n^2}$ both have $r = 1$, while the first one (harmonic series) diverges, but the second one converges. □

**Theorem 25** (Root Test)**.** *Suppose* $\sum_{n=0}^{\infty} a_n$ *is a series of complex numbers, and set*

$$r = \lim_{n\to\infty} \sqrt[n]{|a_n|}.$$

*Then*

*a) If* $r < 1$, $\sum_{n=0}^{\infty} a_n$ *converges absolutely.*

*b) If* $r > 1$, $\sum_{n=0}^{\infty} a_n$ *diverges.*

*c) If* $r = 1$, *the test is inconclusive.*

*Proof.* If $r < 1$, we can choose $\lambda$ with $r < \lambda < 1$ and an integer $N$ so that

$$\sqrt[n]{|a_n|} < \lambda \quad \text{for all} \quad n \geq N.$$

Equivalently

$$|a_n| < \lambda^n \quad \text{for all} \quad n \geq N.$$

Since $0 < \lambda < 1$, $\sum_{n=0}^{\infty} \lambda^n$ converges, and the convergence of the series follows from the comparison test. When $r > 1$, choose any number $\lambda$ with $1 < r < \lambda$, and observe that $\sqrt[n]{|a_n|} > \lambda$ for infinitely many indices $n$ by the definition of "lim sup." Therefore, $|a_n| > \lambda^n \to \infty$ as $n \to \infty$ through some subsequence. In particular, the general term of the series $\sum_{n=0}^{\infty} a_n$ does not tend to zero; therefore the series is divergent. □

**Remark 15.** The root test is more powerful and has wider scope than the ratio test. However, the ratio test is easier to apply, for example, for series involving factorials. Both tests conclude divergence from the fact that $a_n$ does not tend to zero as $n \to \infty$.

## Exercises

1. Suppose $\{x_n\}$ and $\{y_n\}$ are sequences such that $x_n \to L$ and $y_n \to M$. Prove that:

a) $(x_n + y_n) \to L + M$.

b) $cx_n \to cL$ for any constant $c$.

2. Consider the following sequences and using the definition of limit to show that:

a) $\lim_{n\to\infty} \frac{1}{n} = 0$.

b) $\lim_{n\to\infty} \frac{n}{n+1} = 1$.

c) $\lim_{n\to\infty} \frac{n}{2^n} = 0$.

3. Let $f : C \to \mathbb{R}$ be a map, where C is a countable set (recall that C is countable if and only if there exists a bijective mapping $g : \mathbb{N} \to C$). Prove that we can express $f(C) = a_k$, where $a_k \in \mathbb{R}$ and $k \in \mathbb{N}$.

4. Show that if $\{x_n\}$ converges to $L$, then $\{|x_n|\}$ converges to $|L|$. What about the converse?

5. Let $S$ be a nonempty subset of $\mathbb{R}$ which is bounded above. Let $s = \sup S$. Show that there exists a sequence $\{x_n\}$ in $S$ which converges to $s$.

6. Show that $\{x_n\}$ defined by

$$x_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n}$$

is divergent.

7. Show that $\{x_n\}$ defined by

$$x_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n} - \ln(n)$$

is convergent.

8. If $\{x_n\}$ and $\{y_n\}$ are Cauchy sequences in $\mathbb{R}$, then show that $\{x_n + y_n\}$ and $\{x_n y_n\}$ are Cauchy sequences in $\mathbb{R}$.
9. Show that the sequence $\{x_n\}$ defined by

$$x_n = \int_1^n \frac{\cos(t)}{t^2} dt$$

is Cauchy.

10. Let $\{x_n\}$ be a sequence such that there exist $A > 0$ and $C \in (0, 1)$ for which

$$|x_{n+1} - x_n| \leq AC^n$$

for any $n \geq 1$. Show that $\{x_n\}$ is Cauchy. Is this conclusion still valid if we assume only

$$\lim_{n\to\infty} |x_{n+1} - x_n| = 0 \ ?$$

11. Show that if a subsequence $\{x_{n_k}\}$ of a Cauchy sequence $\{x_n\}$ is convergent, then $\{x_n\}$ is convergent.

12. Let $\{x_n\}$ be defined by

$$x_1 = 1 \ \text{ and } \ x_{n+1} = \frac{1}{2}\left(x_n + \frac{2}{x_n}\right) \ .$$

Show that $\{x_n\}$ is convergent and find its limit.

13. Find the limit superior and limit inferior for the following sequences $\{x_n\}$:

a) $x_n = 2^n$.

b) $x_n = n$.

c) $x_n = 1 + (-1)^n + 1/2^n$.

14. If $\{x_n\}$ and $\{y_n\}$ are bounded real sequences, show that

$$\limsup_{n\to\infty}(x_n + y_n) \leq \limsup_{n\to\infty} x_n + \limsup_{n\to\infty} y_n.$$

15. Suppose that $\displaystyle\sum_{n=0}^{\infty} x_n$ is a series of positive terms which is convergent. Show that $\displaystyle\sum_{n=0}^{\infty} \frac{1}{x_n}$ is divergent. What about the converse?

16. Find the sum of the series

a) $\displaystyle\sum_{k=1}^{\infty} \frac{1}{k^2+k}$.

b) $\displaystyle\sum_{k=1}^{\infty} \frac{\ln\left(\dfrac{k^{k+1}}{(k+1)^k}\right)}{k(k+1)}$.

17. Suppose the ratio test applies to $\displaystyle\sum_{n=1}^{\infty} a_n$, i.e., $r = \lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right|$ exists. Show that $\limsup_{n\to\infty} |a_n|^{1/n} = r$.

## 1.5 Topology of the Real Line

Given a real number $a$ and an $\epsilon > 0$, the *$\epsilon$-neighborhood* of the point $a$ is the set

$$V_\epsilon(a) = \{x \in \mathbb{R} : \quad |x-a| < \epsilon\}.$$

In other words $V_\epsilon(a) = (a-\epsilon, a+\epsilon)$ is an open interval centered at $a$ of radius $\epsilon$.

**Definition 18.** The set $O \subset \mathbb{R}$ is *open* if for all points $x \in O$ there exists an $\epsilon$-neighborhood $V_\epsilon(x) \subset O$.

**Example 31.**

a) Any open interval of the form $(a,b) = \{x \in \mathbb{R} : \quad a < x < b\}$ is an open set. Let $x \in (a,b)$ be an arbitrary point. If we take $\epsilon = \min\{x-a, b-x\}$, then it follows that $V_\epsilon(x) \subset (a,b)$. Note that this argument does not work if the interval under consideration is either $[a,b)$ or $(a,b]$ (Figure 1.20).

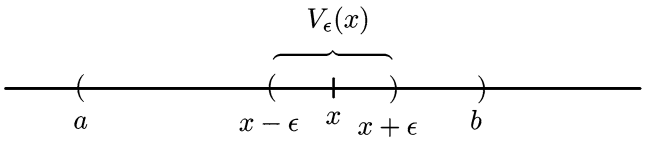



Figure 1.20: The $\epsilon$-neighborhood of the point $x$

b) The set $\mathbb{R}$ is an open set. Given any point $x \in \mathbb{R}$, we can select any $\epsilon$-neighborhood we like and $V_\epsilon(x) \subset \mathbb{R}$ always holds true.

c) From the definition of an open set, we also see that the empty set $\emptyset$ is open in a trivial sense.

The union of two open intervals is again an open set, in fact we have the following theorem:

**Theorem 26.**

*a) The union of an arbitrary collection of open sets is open.*

*b) The intersection of a finite collection of open sets is open.*

*Proof.*

a) Let $\{O_i\}_{i\in I}$ be a collection of open sets in $\mathbb{R}$, and let $O = \bigcup_{i\in I} O_i$. Let $x$ be an arbitrary element of $O$. Since $O$ is the union of $O_i$'s, $x \in O$ implies that there is at least one particular $O_i$, where $x \in O_i$. Because we are assuming $O_i$ is open, there is an $\epsilon$-neighborhood $V_\epsilon(x)$ of $x$ such that $V_\epsilon(x) \subset O_i$. The fact $O_i \subset O$ implies that $V_\epsilon(x) \subset O$ as well.

b) Let $\{O_i\}_{i=1}^N$ be a finite collection of open subsets of $\mathbb{R}$. Take $x \in \bigcap_{i=1}^N O_i$, then $x \in O_i$ for each $1 \leq i \leq N$. By the definition of open set, we know that for each $1 \leq i \leq N$, there exists an $\epsilon_i$-neighborhood of $x$ such that $V_{\epsilon_i}(x) \subset O_i$. Now we seek a single $\epsilon$-neighborhood of $x$ contained in every $O_i$. The natural candidate for $\epsilon$ is $\epsilon = \min\{\epsilon_1, \epsilon_2, \ldots, \epsilon_N\}$. It then follows that

$$V_\epsilon(x) \subseteq \bigcap_{i=1}^N O_i$$

as claimed. □

**Remark 16.** Statement b) above is not true if arbitrary collections are used in place of finite collections, for example, for $i \in \mathbb{N}$, the intervals $\left(\frac{-1}{i}, \frac{1}{i}\right)$ are all open sets, but $\bigcap_{i\in\mathbb{N}} \left(\frac{-1}{i}, \frac{1}{i}\right) = \{0\}$ and $\{0\}$ is not open. We will see that it indeed is a closed set.

We now give a description of the structure of open sets in $\mathbb{R}$.

**Theorem 27.** *Every open set $O$ in $\mathbb{R}$ can be written uniquely as a countable union of disjoint open intervals.*

*Proof.* Let $O$ be an open set in $\mathbb{R}$ and $x \in O$. Since $O$ is an open set, there is a non-trivial open interval (we called this a "neighborhood" of $x$ previously) $x$ is contained in. Next define two points $a_x$ and $b_x$ as $a_x = \inf\{a < x : (a, x) \subset O\}$ and $b_x = \sup\{b > x : (x, b) \subset O\}$, and denote the interval containing $x$ as $I_x = (a_x, b_x)$. This way every $x \in O$ is in $I_x$ and moreover $I_x \subset O$. Thus $O = \bigcup_{x\in O} I_x$. We still need to show this union of intervals $\{I_x\}$ is disjoint and countable. Suppose it is not disjoint. Then we have two open intervals $I_x$ and $I_y$ such that $I_x \cap I_y \neq \emptyset$. Consider $I_x \cup I_y$, which is an open set contained in $O$ containing $x$. However by construction $I_x$ is the maximal one satisfying this

property, which forces $(I_x \cup I_y) \subset I_x$, similarly $(I_x \cup I_y) \subset I_y$, which in turn implies that $I_x = I_y$, meaning any distinct intervals in the collection must be disjoint. To prove that the union $O = \bigcup_{x \in O} I_x$ is countable, we need to recall that the set of rational numbers is countable and every interval $I_x$ contains a rational number. Since different intervals are disjoint, they must contain distinct rationals, and thus the above union is a countable union. □

**Remark 17.** Because the representation of $O = \bigcup_{x \in O} I_i$ is unique and the union is disjoint, the above theorem enables us to "measure" the open set $O$ in terms of the sum of the length of the intervals $I_i$, i.e., the measure of $O = \sum_{i=1}^{\infty} |I_i|$. This is also the beginning of the concept of "measure." However this fact is about open sets, it does not extend to other sets in $\mathbb{R}$, and there is no direct analog of this theorem for subsets of $\mathbb{R}^n$.

**Definition 19.** Let $A \subset \mathbb{R}$. A point $x \in A$ is called an *interior* point of $A$ if there is an open set $U$ such that $x \in U \subset A$. The interior of $A$ is the collection of all interior points of $A$ and is denoted by $\mathring{A}$. This set might be empty.

The condition $x \in \mathring{A}$ is equivalent to the following: There is $\epsilon > 0$ such that $(x - \epsilon, x + \epsilon) \subset A$. If $A = \{a\}$ is a single point set, then $\mathring{A} = \emptyset$, and if $A = [a, b]$, then $\mathring{A} = (a, b)$. The interior of a set can also be described as the union of all open subsets of $A$; thus $\mathring{A}$ is the largest open subset of $A$. It is clear that $A$ is open if and only if $A = \mathring{A}$.

**Definition 20.** Let $S$ be a subset of $\mathbb{R}$. We say $S$ is a *closed* set in $\mathbb{R}$ if the complement of $S$ is an open set in $\mathbb{R}$.

**Example 32.**

a) The empty set $\emptyset$ and $\mathbb{R}$ are both open and closed subsets of $\mathbb{R}$.

b) A single point set $\{a\}$ is closed because its complement is the union of two open sets, i.e., its complement is $(-\infty, a) \cup (a, +\infty)$.

c) An interval of the form $[a, b]$ is a closed set.

**Theorem 28.**

*a) The intersection of an arbitrary collection of closed sets is closed.*

*b) The union of a finite collection of closed sets is closed.*

*Proof.*

a) Let $\{S_i\}_{i \in I}$ be any collection of closed subsets of $\mathbb{R}$. The proof follows from De Morgan's law:

$$\left( \bigcap_{i \in I} S_i \right)^c = \bigcup_{i \in I} S_i^c$$

and the first part of Theorem 26.

b) If $\{S_i\}_{i=1}^N$ is a finite collection of closed subsets of $\mathbb{R}$, then by De Morgan's law we have

$$\left(\bigcup_{i=1}^{N} S_i\right)^c = \bigcap_{i=1}^{N} S_i^c.$$

Now apply the second part of Theorem 26 to complete the proof. □

**Remark 18.** Statement b) above is not true if arbitrary collections are used in place of finite collections; for example, if we take closed sets $S_k = \left[\frac{1}{k+1}, \frac{k}{k+1}\right]$, then $\bigcup_{k\in\mathbb{N}} \left[\frac{1}{k+1}, \frac{k}{k+1}\right] = (0,1)$.

Theorems 26 and 28 have many applications. In particular they allow us to define the largest open set contained in a given set $S$ and the smallest closed set containing the given set $S$ as in the following:

**Definition 21.** Let $S$ be a subset of $\mathbb{R}$.

a) The *interior* of $S$ is the set

$$\mathring{S} := \bigcup \{O \subseteq S \text{ and } O \text{ is open in } \mathbb{R}\}.$$

b) The *closure* of $S$ is the set

$$\overline{S} := \bigcap \{F \supseteq S \text{ and } F \text{ is closed in } \mathbb{R}\}.$$

Note that every set $S$ contains the empty set $\emptyset$ and is contained in $\mathbb{R}$; hence $\overline{S}$ and $\mathring{S}$ are well defined. Moreover, by Theorems 26 and 28 we immediately see that $\overline{S}$ is a closed set and $\mathring{S}$ is an open set. The following result claims more: $\overline{S}$ is the smallest closed set which contains $S$ and $\mathring{S}$ is the largest open set contained in $S$.

**Proposition 4.** *Let $S$ be a subset of $\mathbb{R}$. Then*

*a) $\mathring{S} \subseteq S \subseteq \overline{S}$.*

*b) If $G$ is open and $G \subseteq S$, then $G \subseteq \mathring{S}$.*

*c) If $K$ is closed and $K \supseteq S$, then $K \supseteq \overline{S}$.*

*Proof.* The proof of part a) is clear. For the proof of parts b) and c), use the above definition, and observe that if $O$ is an open set contained in $S$, then $O \subseteq \mathring{S}$ and $K$ is a closed set containing $S$, and then $\overline{S} \subseteq K$. □

**Remark 19.** Parts b) and c) imply that

$$S = \mathring{S} \quad \text{if and only if} \quad S \text{ is open}$$

and

$$S = \overline{S} \quad \text{if and only if} \quad S \text{ is closed.}$$

Furthermore, these concepts have a natural extension to the subsets of $\mathbb{R}^n$ or to an arbitrary metric space $M$ which we will define in Chapter 2. For a detailed discussion of the topology of $\mathbb{R}^n$ or an arbitrary metric space, we refer the reader to [52], pp. 303–342.

**Definition 22.** Let $S$ be a subset of $\mathbb{R}$. We say $x \in \mathbb{R}$ is an *accumulation point* of $S$ if, for every $\epsilon > 0$, we have $((x-\epsilon, x+\epsilon) \setminus \{x\}) \cap S \neq \emptyset$.

Thus, $x$ is an accumulation point of $S$ if every interval around $x$ contains points of $S$ other than $x$. A set need not have any accumulation points, for example, a set consisting of a single point or the set of integers in $\mathbb{R}$ has no accumulation points. Moreover, the set of accumulation points of a set need not lie in that set. For example, if we set $S = (0, 1)$, then the set of accumulation points of S is the whole interval $[0, 1]$. The following lemma clarifies the concept of accumulation point further.

**Lemma 6.** *Let $S \subset \mathbb{R}$. Then $x \in \mathbb{R}$ is an accumulation point of $S$ if and only if every neighborhood of $x$ contains infinitely many points of $S$.*

*Proof.* Let $x$ be an accumulation point of $S$ and $\epsilon > 0$. By Definition 22 there is an $\epsilon$-neighborhood of $x$, $(x-\epsilon, x+\epsilon)$ such that $x_1 \in S \cap (x-\epsilon, x+\epsilon)$ and $x_1 \neq x$. Let $\epsilon_1 = |x - x_1| > 0$. Again using the definition of an accumulation point, we can find $x_2 \in S \cap (x-\epsilon_1, x+\epsilon_1)$ such that $x_2 \neq x$. Continuing this process will yield an infinite set of elements in $S \cap (x-\epsilon, x+\epsilon)$. □

Let $S^*$ denote a set of all accumulation points of $S$. If $x \in S$ but $x \notin S^*$, then $x$ is called an *isolated* point of $S$. Note that an isolated point is always an element of the set, while an accumulation point does not necessarily belong to the set. Accumulation points are also referred to as "cluster points" or "limit points."

**Proposition 5.** *A point $x$ is an accumulation point of a set $S$ if and only if $x = \lim_{n\to\infty} x_n$ for some sequence $\{x_n\}$ contained in $S$ satisfying $x_n \neq x$ for all $n \in \mathbb{N}$.*

*Proof.* Given a point $s \in S$, where $s$ is the limit of the constant sequence $\{s, s, s, \dots\}$. To handle this uninteresting situation, the condition $x_n \neq x$ is given in the statement of the theorem. Assume $x$ to be an accumulation point of the set $S$, and we want to produce a sequence $\{x_n\}$ with $x_n \to x$. Using the definition of accumulation point and by taking $\epsilon_n = \dfrac{1}{n}$, we can find

$$x_n \in V_{\frac{1}{n}}(x) \cap S \quad \text{for each} \quad n \in \mathbb{N}$$

with the condition that $x_n \neq x$. Given arbitrary $\epsilon > 0$, choose $N$ so that $\frac{1}{N} < \epsilon$, and then we have $|x_n - x| < \epsilon$ holds true for all $n \geq N$. We leave the proof of the reverse implication to the reader. □

The adjective *closed* is often used in mathematics; it roughly means if an operation on the elements of a given set is performed, we still obtain an element of the same given set. For example, for the definition of a vector space, one can simply state a vector space is a set which is closed under addition and scalar multiplication. However, in analysis, the operation one considers is the limiting operation. The following proposition follows directly from the definition of a closed set.

**Proposition 6.** *A set $S \subseteq \mathbb{R}$ is closed if it contains all of its accumulation points.*

**Example 33.** To prove the closed interval $[a, b] = \{x \in \mathbb{R} : \quad a \leq x \leq b\}$ is a closed set, let $x$ be an accumulation point of $[a, b]$. Then there exists a sequence $\{x_n\}$ such that $x_n \to x$. We need to show that $x \in [a, b]$. Using the relationship between inequalities and the limit, we obtain

$$a \leq x_n \leq b \quad \text{implies that} \quad a \leq x \leq b,$$

proving $x \in [a, b]$ and the set $[a, b]$ is closed.

**Definition 23.** Given $S \subseteq \mathbb{R}$ and $S^*$ the set of all accumulation points of $S$, the *closure* of a set $S$ is defined to be $\overline{S} = S \cup S^*$.

**Example 34.**

a) If $S = (a, b)$, then $\overline{S} = [a, b]$.

b) If $S$ is a closed interval $[a, b]$, then $\overline{S} = S$.

c) If $S = \left\{ \frac{1}{n} : \ n \in \mathbb{N} \right\}$, then the closure of $S$ is $\overline{S} = S \cup \{0\}$.

d) If the set $S$ equals to $\mathbb{Q}$, the set of rational numbers and $y \in \mathbb{R}$ an arbitrary element, the density of rationals in $\mathbb{R}$ allows us to find a rational number $r$, where $r$ is in any $\epsilon$-neighborhood of $y$ with $y \neq r$. Thus, $y$ is an accumulation point of $\mathbb{Q}$ and therefore $\overline{\mathbb{Q}} = \mathbb{R}$.

**Definition 24.** Let $S \subseteq \mathbb{R}$. The *boundary* of $S$ is the set

$$\partial S := \{x \in \mathbb{R} : \text{ for all } \epsilon > 0, \ V_\epsilon(x) \cap S \neq \emptyset \text{ and } V_\epsilon(x) \cap S^c \neq \emptyset\}.$$

Let $S \subseteq \mathbb{R}$. Recall that a point $x \in \mathbb{R}$ is an *interior point* of $S$ if there exists an $\epsilon$-neighborhood $V_\epsilon(x)$ of $x$ such that $V_\epsilon(x) \subset S$. The set of all interior points of $S$ is denoted by $\mathring{S}$. For the following sets, we find their interior and boundary points.

**Example 35.**

a) Let $S_1 = (0,1)$, $S_2 = (0,1]$, and $S_3 = [0,1]$. Then

$$\mathring{S_1} = \mathring{S_2} = \mathring{S_3} = (0,1) \quad \text{and} \quad \partial S_1 = \partial S_2 = \partial S_3 = \{0,1\}.$$

b) Let $S = (0,1) \cup (1,2]$, then $\partial S = \{0,1,2\}$ and $\mathring{S} = (0,1) \cup (1,2)$.

c) If $S = \mathbb{R}$, then $\partial S = \emptyset$ and $\mathring{S} = S$.

d) If $S = \{x \in \mathbb{R} : \ x \in [0,1] \text{ and } x \text{ is rational}\}$, then $\partial S = [0,1]$, since for any $\epsilon > 0$ and $x \in [0,1]$, $V_\epsilon(x) = (x - \epsilon, x + \epsilon)$ contains both rational and irrational points.

Some basic relationships between accumulation points, closure, and closed sets are given in the following theorem.

**Theorem 29.** *Let $S$ be a subset of $\mathbb{R}$. Then*

*a) $S$ is a closed set if and only if $S$ contains all of its accumulation points.*

*b) $\overline{S}$ is a closed set.*

*c) $S$ is closed if and only if $S = \overline{S}$.*

*d) $\overline{S} = S \cup \partial(S)$.*

*Proof.* To prove a) assume that $S$ is closed, and let $S^*$ denote the set of all accumulation points of $S$. We want to show $S^* \subseteq S$. Let $x \in S^*$. Since $S$ is closed, $\mathbb{R} \setminus S$ is an open set; thus if $x \in \mathbb{R} \setminus S$, there is an $\epsilon > 0$ such that $V_\epsilon(x) \subset \mathbb{R} \setminus S$, i.e., $V_\epsilon(x) \cap S = \emptyset$. Thus $x$ is not an accumulation point, a contradiction to $x \in S^*$ and so $S$ contains all of its accumulation points. Conversely, suppose $S^* \subseteq S$. We will show that $\mathbb{R} \setminus S$ is an open set. To this end, let $x \in \mathbb{R} \setminus S$. Then $x \notin S$ and thus $x \notin S^*$ (not an accumulation point). There is an $\epsilon$-neighborhood $V_\epsilon(x)$ such that $V_\epsilon(x) \cap S = \emptyset$; i.e., $V_\epsilon(x) \subset \mathbb{R} \setminus S$. Hence $\mathbb{R} \setminus S$ is open, and so $S$ is closed.
To prove part b), we use part a). We must show $\overline{S}^* \subset \overline{S}$, where $\overline{S}^*$ denotes the set of all accumulation points of $\overline{S}$. If $x \in \overline{S}^*$, then every deleted neighborhood $V_\epsilon^{'}(x) = (x-\epsilon, x) \cup (x, x+\epsilon)$ intersects with $\overline{S}$. We must show that $V_\epsilon^{'}(x) \cap S \neq \emptyset$. Let $y \in V_\epsilon^{'}(x) \cap \overline{S}$; since $V_\epsilon^{'}(x)$ is an open set, there exists a neighborhood $V_\delta(y)$ such that $V_\delta(y) \subset V_\epsilon^{'}(x)$. See Figure 1.21.

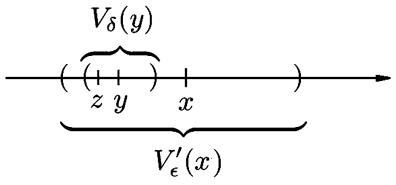



Figure 1.21: Every deleted neighborhood of $x$ intersects with $\overline{S}$

However, $y \in \overline{S}$, so every neighborhood of $y$ intersects with $S$. This means that there is a point $z \in V_\delta(y) \cap S$ with $z \neq y$. But then

$$z \in V_\delta(y) \subseteq V_\epsilon^{'}(x),$$

so that $x \in S^*$ and $x \in \overline{S}$.
Proofs of parts c) and d) are left to the reader. □

It turns out that the boundary of set can be described using the difference of two familiar sets, as stated in the next theorem.

**Theorem 30.** *Let $S$ be a subset of $\mathbb{R}$. Then $\partial S = \overline{S} \setminus \mathring{S}$.*

*Proof.* Using the definition of the boundary of set, we must show that

$$x \in \overline{S} \quad \text{if and only if} \quad V_r(x) \cap S \neq \emptyset \quad \text{for all } r > 0$$

and

$$x \notin \mathring{S} \quad \text{if and only if} \quad V_r(x) \cap S^c \neq \emptyset \quad \text{for all } r > 0.$$

We will leave the details to the reader. □

## Compact Sets

Compactness is an important concept in the analysis. It is what reduces the infinite to the finite. There are a number of equivalent ways to describe compactness in $\mathbb{R}$, and we can use whatever description of compactness is most appropriate for a given situation. Intuitively, a compact set in $\mathbb{R}$ is closed and contained in a bounded region. This is called the Heine–Borel Theorem which has a number of profound and far reaching applications in the analysis. For example, if a function is defined on a compact set, then it is automatically bounded and uniformly continuous. See Theorem 42 in the next section.

**Definition 25.** Let $S \subseteq \mathbb{R}$. An *open cover* for $S$ is a collection of open sets $\{O_\alpha : \alpha \in \Gamma\}$ whose union contains $S$, that is, $S \subset \bigcup_{\alpha \in \Gamma} O_\alpha$. A *finite subcover* is a finite subcollection $\{O_{\alpha_i} : \alpha_i \in \Gamma,\ i = 1, 2, \ldots, m\}$ that covers $S$.

For example, the collection of open sets $\left\{\left(\frac{-1}{k}, \frac{k+1}{k}\right)\right\}$ is a covering for the interval $(0, 1)$. An open cover for a set may or may not have a finite subcover. If *every* open cover for a set has a finite subcover, then the set is called a compact set.

**Definition 26.** A subset $S$ of $\mathbb{R}$ is called *compact* if every open cover for $S$ contains a finite subcover.

To understand this definition, let us start with some examples.

**Example 36.**

a) The empty set and all finite subsets of $\mathbb{R}$ are compact. The empty set needs no set to cover, and any finite set $F \subset S$ can be covered by finitely many open sets, one for each element of $F$.

b) The set $S = (0, 1)$ is not compact, because the open cover $\{O_n = \left(\frac{1}{n}, 2\right)\}$ has no finite subcover. To see this let $0 < x < 1$; then by the Archimedean property, there exists $p \in \mathbb{N}$ such that $\frac{1}{p} < x$. Thus $x \in O_p$ and $\{O_n = \left(\frac{1}{n}, 2\right)\}$ is an open cover for $(0, 1)$. Now consider a finite subcollection, $\{O_{n_1}, O_{n_2}, \ldots, O_{n_k}\}$, and set $m = \max\{n_1, n_2, \ldots, n_k\}$; then

$$O_m = O_{n_1} \cup O_{n_2} \cup \cdots \cup O_{n_k} = \left(\frac{1}{m}, 2\right).$$

This finite subcollection cannot cover $(0, 1)$. Since we found a particular cover with no finite subcover, we conclude that $(0, 1)$ is not compact. See Figure 1.22.

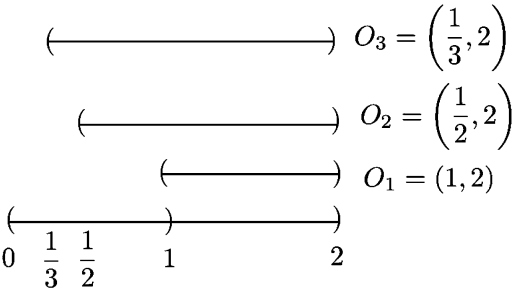



Figure 1.22: $S = (0, 1)$ is not compact

c) The real line $\mathbb{R}$ is not compact, because the open cover $\{(n - 1, n + 1) : n = 0, \pm 1, \pm 2, \ldots\}$ for $\mathbb{R}$ has no finite subcover. Why?

d) The interval $[0, \infty)$ is not compact, because the open cover $\{(-1, n)\}$ has no finite subcover. Why?

## Properties of Compact Sets

**Theorem 31.** *The following are basic properties of compact sets:*

*a) Every compact set is closed.*

*b) A closed subset of a compact set is compact.*

*c) If $F$ is closed and $S$ is compact, then $F \cap S$ is compact.*

*d) If $S$ is a nonempty, closed, and bounded subset of $\mathbb{R}$, then $S$ has a maximum and a minimum.*

*Proof.*

a) Let $S$ be a compact subset of $\mathbb{R}$; we shall prove that the complement of $S$ in $\mathbb{R}$ is open. Let $x \in \mathbb{R}$, $x \notin S$. If $y \in S$, let $r < |x - y|$ and $V_x$ and $W_y$ be neighborhoods of $x$ and $y$, respectively, of radius less than $r$. Since $S$ is compact, there are finitely many points $y_1, y_2, \ldots, y_n$ in $S$ such that

$$S \subset W_{y_1} \cup W_{y_2} \cup \cdots \cup W_{y_n} = W.$$

If $V = V_{x_1} \cap V_{x_2} \cap \cdots \cap V_{x_n}$, then $V$ is a neighborhood of $x$ which does not intersect with $W$. Hence $V$ is contained in the complement of $S$, $V \subset \mathbb{R} \setminus S = S^c$, $S^c$ is open, and thus $S$ is closed.

b) Let $F$ be a closed subset of $S$, where $S$ is compact in $\mathbb{R}$. Let $\{O_\alpha : \alpha \in \Gamma\}$ be an open covering for $F$. Now $\mathbb{R} \setminus F = F^c$ is open; hence $\{F^c\} \cup \{O_\alpha : \alpha \in \Gamma\}$ is an open covering of $S$. Since $S$ is compact, there is a finite subcover for $S$

$$S \subseteq F^c \cup \left( \bigcup_{i=1}^{n} O_{\alpha_i} \right).$$

But $F^c \cap F = \emptyset$, so we may remove it from the above union and still retain an open cover for $F$ with a finite subcover.

c) Since compact sets are closed and the intersection of two closed sets is closed, $F \cap S$ is closed and $F \cap S \subset S$; thus by item b) above $F \cap S$ is compact.

d) Since $S$ is nonempty and bounded above, we can appeal to the completeness axiom to conclude $m = \sup S$ exists. Using the definition of the supremum, we know that given $\epsilon > 0$, $m - \epsilon$ is not an upper bound for $S$. Next, we claim $m \in S$; thus $m = \max S$. If $m \notin S$, then there exists $x \in S$ such that $m - \epsilon < x < m$; this means $m$ is an accumulation point of $S$. But $S$ is closed and thus contains all of its accumulation points. It must be that $m \in S$. A similar argument can be given for $\inf S$ to conclude $\inf S = \min S \in S$. □

It is often difficult to show directly that a given set satisfies the definition of compactness. The definition of a compact set requires us to consider *every* open covering and show that it has finite subcover. We cannot replace *every* by *some* open cover either. Consider $S = \mathbb{R}$ itself, and $\mathbb{R}$ is not compact, but it is an open set and $\mathbb{R} \subseteq \mathbb{R}$ (in which case an open cover is the same as finite subcover). Thus the definition of compact set is useful in showing a set is not compact. Then all we need is to find some open cover for the set which does not have a finite subcover as we have done in the above examples. Fortunately, the *Heine–Borel Theorem* gives us a much easier characterization of compact subsets of $\mathbb{R}^n$. Although this theorem is valid for $\mathbb{R}^n$, we will state the theorem and give a proof for $\mathbb{R}$.

**Theorem 32** (Heine–Borel Theorem)**.** *A subset $S$ of $\mathbb{R}$ is compact if and only if it is closed and bounded.*

*Proof.* $\Rightarrow$): Suppose $S$ is compact, and let $\{O_n = (-n, n)\}$ be an open cover for $S$. Since $S$ is compact, it has a finite subcover, i.e., $S \subset \bigcup_{j=1}^{k}(-n_j, n_j)$. Setting $n_0 = \max\{n_1, n_2, \ldots, n_k\}$, we get $S \subset (-n_0, n_0)$; thus $S$ is bounded. Suppose $S$ is compact but not closed, i.e., $S \neq \overline{S} := S \cup \{\text{accumulation points of } S\}$. Let $x_0 \in \overline{S} \setminus S$; then $x_0$ is an accumulation point of $S$, and thus for $\epsilon > 0$ there is a neighborhood $V_\epsilon(x_0) = (x_0 - \epsilon, x_0 + \epsilon)$, which contains $y \in S$ and $y \neq x$. For each $n \in \mathbb{N}$, we let $U_n = \mathbb{R} \setminus [x_0 - \frac{1}{n}, x_0 + \frac{1}{n}]$. Now each $U_n$ is an open set, and we have

$$S \subseteq \bigcup_{n=1}^{\infty} \mathbb{R} \setminus \left[x_0 - \frac{1}{n}, x_0 + \frac{1}{n}\right] = \mathbb{R} \setminus \{x_0\}.$$

But $S$ is compact; thus there exist $n_1 < n_2 < \cdots < n_k$ in $\mathbb{N}$ such that

$$S \subseteq \bigcup_{j=1}^{k} U_{n_j} = \mathbb{R} \setminus \left[x_0 - \frac{1}{n_0}, x_0 + \frac{1}{n_0}\right],$$

where $n_0 = \max\{n_1, n_2, \ldots, n_k\}$, which implies that $S \cap [x_0 - \frac{1}{n_0}, x_0 + \frac{1}{n_0}] = \emptyset$, contradicting our choice of $x_0 \in \overline{S} \setminus S$.

$\Leftarrow$): Conversely, suppose $S$ is a closed and bounded subset of $\mathbb{R}$. If $a = \sup S$ and $b = \inf S$, then since $S$ is closed, $a$ and $b$ are in both $S$ and $S \subseteq [a, b]$. Let $\{U_i\}_{i \in I}$ be an open cover for $S$. By adding the complement of $S$ to this collection, we obtain a collection $\mathcal{U}$ of open sets whose union covers $[a, b]$, i.e., $[a, b] \subseteq \left(\bigcup U_i\right) \cup S^c$. Let

$$F = \{x \in [a, b] : \ [a, x] \text{ is covered by the finite number of open sets in } \mathcal{U}\}.$$

Since $a \in F$, $F \neq \emptyset$ and $F$ is bounded above by $b$. If we let $c = \sup F$, then $c \in F$ since $F$ is closed. We want to show $c = b$. If $c \neq b$, there exists $y$ such that $c < y < b$, and by letting $\mathcal{U}^*$ be a finite sub-collection of $\mathcal{U}$ that covers $[a, c]$, $[c, y]$ is contained in an open set in $\mathcal{U}^*$ that contains $c$. Thus $[a, y]$ is covered by $\mathcal{U}^*$. This is a contradiction; hence we must have $b = c$. By throwing away the complement of $S$, we obtain a finite cover for $S$ from the collection $\mathcal{U}$. □

The following examples illustrate how useful this theorem is.

**Example 37.**

a) The set $S = [0, 1] \cup [4, 5]$ is compact because it is both closed and bounded.

b) The set $S = \{x \in \mathbb{R} : \ x \geq 0\}$ is not compact because it is not bounded.

c) The set $S = [1, 2] \cap \mathbb{Q}^c$ is not compact because $S$ is not closed. (Hint: Consider $3/2 \notin S$ and a neighborhood of $3/2$ to show that $S^c$ is not open).

In the section on Sequences and Series, (see Theorem 15), we have seen the proof of the *Bolzano–Weierstrass property* which states:

"Every bounded sequence of real numbers has a convergent subsequence."

Now consider a sequence $\{x_n\} \subset S$, where $S$ is a compact set. Since compact sets are bounded, then $\{x_n\}$ is bounded, and by the Bolzano–Weierstrass property $\{x_n\}$ has a subsequence $\{x_{n_k}\}$ that converges to some point $x$. But $S$ is also closed; thus $S$ contains the limit point. This gives us the following definition of compactness, sometimes referred to as *sequential compactness.*

**Definition 27** (Sequential Compactness)**.** A set $S$ is called *sequentially compact* in $\mathbb{R}$ if every sequence in $S$ has a subsequence that converges to a limit that is also in $S$.

**Theorem 33.** *Let $S$ be a subset of $\mathbb{R}$. Then the following are equivalent:*

*a) Any open cover for $S$ has a finite subcover.*

*b) $S$ is closed and bounded.*

*c) Every sequence in $S$ has a convergent subsequence that converges to a limit in $S$.*

*Proof.* Equivalence of a) and b) is the Heine–Borel Theorem. Equivalence of a) and c) is called the Bolzano–Weierstrass Theorem , and we already gave an idea why closed and bounded implies sequential compactness. We need to show sequential compactness implies closed and bounded. Assume c) is true. If $S$ is unbounded, then $S$ contains points $x_n$ with $|x_n| > n$ for $n = 1, 2, \ldots$. The set containing these points $x_n$ is infinite and clearly has no limit point in $\mathbb{R}$ and hence none in $S$. This sequence could not have a convergent subsequence. Thus c) implies $S$ is bounded. If $S$ is not closed, then there is a point $x_0 \in \mathbb{R}$ which is a limit point of the subsequence but not a point of $S$. For $n = 1, 2, \ldots$, there are points $x_{n_k}$ such that $|x_{n_k} - x_0| < 1/n$. Let $E$ be the set of these points $x_{n_k}$. Then $E$ is an infinite set and E has $x_0$ as a limit point. If $y \in \mathbb{R}$ with $y \neq x_0$, then

$$|x_{n_k} - y| \geq |x_0 - y| - |x_{n_k} - x_0| > |x_0 - y| - \frac{1}{n} \geq \frac{1}{2}|x_0 - y|$$

for all but finitely many $n$; thus $y$ is not a limit point of $E$ or $\{x_{n_k}\}$ has no limit point in $S$; hence $S$ must be closed. □

**Remark 20.** The equivalences given in the above theorem are still true if we replace $\mathbb{R}$ by $\mathbb{R}^n$. The equivalence of a) and c) still holds in metric spaces; however as we will see later that for an arbitrary metric space b) does not imply a) nor c). See Example 64 in Section 2.1.

**Proposition 7** (Finite Intersection Property)**.** *Let $\mathcal{U} = \{K_\alpha : \alpha \in \Gamma\}$ be a family of compact subsets of $\mathbb{R}$. Suppose that the intersection of any finite subfamily of $\mathcal{U}$ is nonempty. Then $\bigcap\{K_\alpha : \alpha \in \Gamma\} \neq \emptyset$.*

*Proof.* For each $\alpha \in \Gamma$, set $O_\alpha = \mathbb{R} \setminus K_\alpha$ (or equivalently $O_\alpha = K_\alpha^c$), and fix a member $K_1$ from $\mathcal{U}$. Assume that no point of $K_1$ belongs to every $K_\alpha$. Then every point of $K_1$ belongs to some $O_\alpha$. That is, the sets $O_\alpha$ form an open cover for $K_1$; since $K_1$ is compact, there are finitely many indices $\alpha_1, \alpha_2, \ldots, \alpha_n$ such that

$$K_1 \subset O_{\alpha_1} \cup O_{\alpha_2} \cup \cdots \cup O_{\alpha_n}.$$

Using the fact that

$$\mathbb{R} \setminus (K_{\alpha_1} \cap K_{\alpha_2} \cap \cdots \cap K_{\alpha_n}) = (\mathbb{R} \setminus K_{\alpha_1}) \cup (\mathbb{R} \setminus K_{\alpha_2}) \cup \cdots \cup (\mathbb{R} \setminus K_{\alpha_n}),$$

we have

$$K_1 \subset \mathbb{R} \setminus (K_{\alpha_1} \cap K_{\alpha_2} \cap \cdots \cap K_{\alpha_n}),$$

and therefore

$$K_1 \cap (K_{\alpha_1} \cap \cdots \cap K_{\alpha_n}) = \emptyset,$$

in contradiction to our hypothesis. Thus some point in $K_1$ belongs to each $K_\alpha$, and $\bigcap\{K_\alpha : \alpha \in \Gamma\} \neq \emptyset$. □

**Theorem 34** (Nested Set Property for Compact Sets)**.** *If $\{K_n\}$ is a sequence of compact sets in $\mathbb{R}$ such that $K_{n+1} \subset K_n$ for all $n = 1, 2, \ldots$, then there is at least one point in $\bigcap_{n=1}^{\infty} K_n$.*

*Proof.* The result follows from the finite intersection property. The set $K_1$ is compact, and the sets $K_1, K_2, \ldots$ have the finite intersection property. Since the intersection of any finite collection equals $K_n$ with the highest index,

$$K_1 \bigcap \left( \bigcap_{n=1}^{\infty} K_n \right) = \bigcap_{n=1}^{\infty} K_n \neq \emptyset.$$ □

## Connected Sets

We have introduced open and closed sets that are analogs of open and closed intervals in $\mathbb{R}$, and compact subsets of $\mathbb{R}$ are analogs of closed and bounded intervals. Next we define the concept of a *connected* set. Loosely speaking a connected set is one piece and cannot be broken into nonempty open pieces which do not share any common points. For example, the set $A = (0, 1) \cup (1, 2)$ has two pieces, although the two intervals $(0, 1)$ and $(1, 2)$ have a limit point $x = 1$ in common, there is still some "space" between them, meaning no limit point of one of these intervals is actually contained in the other. Said in another way, the closure of $(0, 1)$ does not intersect $(1, 2)$ and vice versa. Moreover, our intuition can fail in judging whether the set

$$A = \left\{ (x, \sin(\frac{1}{x})) : x > 0 \right\} \cup \{(0, y) : y \in [-1, 1]\} \subset \mathbb{R}^2$$

is broken or not. See Figure 1.23.

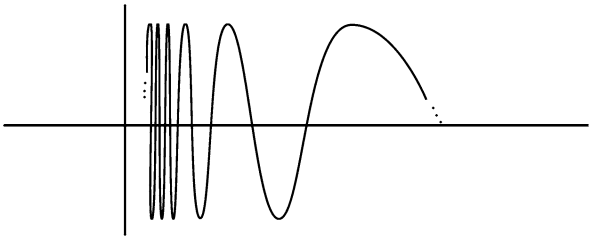


Figure 1.23: Is this set connected?

We must seek a sound mathematical definition that we can depend on. In later sections, we examine continuity and differentiability of functions defined on subsets of $\mathbb{R}$, and there are some very important theorems such as the Intermediate Value Theorem which depend on the fact that an interval is connected. The definition of *connected* is a bit tricky since it is stated as a negation of *disconnected*.

**Definition 28.** A subset $A \subset \mathbb{R}$ is said to be *disconnected* (or not connected) if there exists a pair of open sets $U$ and $V$ in $\mathbb{R}$ such that:

a) $U \cap A \neq \emptyset$, $V \cap A \neq \emptyset$.

b) $U \cap V \cap A = \emptyset$.

c) $A \subset U \cup V$.

A set $A$ is connected if it is not disconnected.

Note that basically two nonempty sets $U$ and $V$ separate $A \subset \mathbb{R}$ if $A = U \cup V$ and $U \cap V = \emptyset$.

**Example 38.**

a) The empty set $\emptyset$ is connected because it can never be written as the union of nonempty sets.

b) The set $A = [-1, 0] \cup [5, 6]$ is disconnected in $\mathbb{R}$ because $U = (-2, 1)$ and $V = (4, 7)$ are disjoint open sets that each intersects with $A$ and jointly contains $A$.

c) The set $\mathbb{Z} \subset \mathbb{R}$ is disconnected. To see this let $U = (1/2, \infty)$ and $V = (-\infty, 1/4)$, then $\mathbb{Z} \subset U \cup V$, $\mathbb{Z} \cap U = \{1, 2, 3, \dots\} \neq \emptyset$, $\mathbb{Z} \cap V = \{\dots, -2, -1, 0\} \neq \emptyset$, and $\mathbb{Z} \cap U \cap V = \emptyset$.

d) The set of rationals $\mathbb{Q}$ is disconnected in $\mathbb{R}$. Because if $r_1, r_2 \in \mathbb{Q}$ with $r_1 < r_2$, we can choose an irrational number $x$ with $r_1 < x < r_2$. Then by taking $r_1 \in U = (-\infty, x)$ and $r_2 \in V = (x, \infty)$, we obtain two disjoint open sets that separate $\mathbb{Q}$.

e) The set $A = \{x \in \mathbb{R} : \ x \neq 0\}$ is also disconnected; in this case $U = (-\infty, 0)$ and $V = (0, \infty)$ does the job.

Note that although the above definition is valid for subsets of $\mathbb{R}^n$ or even for an arbitrary metric space, one needs to pay attention to the metric. For example, $\mathbb{R}$ with the usual metric (absolute value function) is connected, but if we put the *discrete* metric on $\mathbb{R}$, it is not connected. We will see later that in the discrete metric $(-\infty, 0]$ and $(0, \infty)$ are both open sets. From example e) above we also learn that the open sets $U$ and $V$ need not be a positive distance apart. Even a small gap due to omitting the origin creates a disconnected set. The concept of connectedness is more subtle when working with subsets of the plane or higher dimensions. Luckily, connected sets have a simple description in $\mathbb{R}$ which we give in the following theorem.

**Theorem 35.** *A subset $A$ of $\mathbb{R}$ is connected if and only if $A$ is an interval.*

*Proof.* First a clarification: When we say an "interval" what we mean is a set of the form $[a, b], [a, b), (a, b]$, or $(a, b)$, where $a$ and $b$ can be $\mp\infty$ on an open end of the interval. There is another concept of connectedness called *path connectedness* which we will not cover here. But it can be shown that intervals are connected because they are path connected (for details see [18], p. 258). For the converse we assume $A$ is not an interval. This means there are points $x, y$, and $z$ such that $x < y < z$, with $x, z \in A$ and $y \notin A$. Then we can form the open sets $U = (-\infty, y)$ and $V = (y, \infty)$ such that

- $A \subset U \cup V$.
- $U \cap A \neq \emptyset$ and $V \cap A \neq \emptyset$.
- $U \cap \ V \cap A = \emptyset$.

Thus $A$ is disconnected. □

## The Cantor Set

The *Cantor set* or sometimes called the *Cantor ternary set* is an intriguing subset of $\mathbb{R}$ which will extend our understanding of the nature of subsets of $\mathbb{R}$ and is a valuable source of counterexamples in the analysis. For example, the Cantor set is the basis of the construction of a function called "Devil's staircase," which is a continuous, non-decreasing function that is not constant, yet has zero derivative at almost every point. The following construction is due to Georg Cantor, whose name was already mentioned several times, especially in our discussion of uncountable sets. The Cantor set is a subset of $[0, 1]$; it is "large" (uncountable), yet it is somewhat "small" (has length zero).
Consider the process of successively removing "middle thirds" from the interval $[0, 1]$. Let $C_0 = [0, 1]$, and define $C_1$ by removing the open interval $I_1 = \left(\frac{1}{3}, \frac{2}{3}\right)$ from $C_0$, that is,

$$C_1 = C_0 \setminus \left(\frac{1}{3}, \frac{2}{3}\right) = \left[0, \frac{1}{3}\right] \bigcup \left[\frac{2}{3}, 1\right].$$

Now, construct $C_2$ by removing middle thirds from $[0, \frac{1}{3}]$ and $[\frac{2}{3}, 1]$, let $I_2 = (\frac{1}{9}, \frac{2}{9}) \cup (\frac{7}{9}, \frac{8}{9})$, and set

$$C_2 = C_1 \setminus I_2 = \left( \left[0, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{1}{3}\right] \right) \bigcup \left( \left[\frac{2}{3}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, 1\right] \right).$$

Notice that $C_1$ and $C_2$ consist of two and four closed intervals each having length $\frac{1}{3}$ and $\frac{1}{9}$, respectively. If we continue this process inductively, then for each $n = 1, 2, \ldots$ we obtain a set $C_n$ consisting of $2^n$ closed intervals each having length $\dfrac{1}{3^n}$. The Cantor set $C$ is defined as (Figure 1.24)

$$C = \bigcap_{n=0}^{\infty} C_n.$$

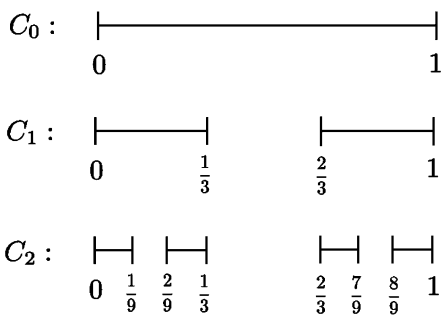



Figure 1.24: Defining the Cantor set $C = \bigcap_{n=0}^{\infty} C_n$

It follows from the nested interval property that $C \neq \emptyset$. Notice that if we continue this process indefinitely, it seems like nothing is left, but if we keep track of the end points of each interval $C_n$, we can see that

$$0, 1, \frac{1}{3}, \frac{2}{3}, \frac{1}{9}, \frac{2}{9}, \cdots \in C.$$

We refer to all of the points in $C$ of the form $\dfrac{a}{3^n}$ for some integers $a$ and $n$ as end points of $C$. We will see that the Cantor set $C$ contains end points and many other points as well. Surprisingly, $C$ is uncountable. Before we prove that $C$ is uncountable, we first prove $C$ is "small":

**Proposition 8.** *The Cantor set $C$ has zero length.*

*Proof.* We start by observing the length of the intervals removed from $[0, 1]$ to form $C$. To form $C_1$ we removed $I_1 = \left(\frac{1}{3}, \frac{2}{3}\right)$ and the length of $I_1 = 1/3$. In the second step we removed $I_2 = \left(\frac{1}{9}, \frac{2}{9}\right) \cup \left(\frac{7}{9}, \frac{8}{9}\right)$ and the length of $I_2 = 2 \cdot (1/9)$.

To construct $C_n$ we removed $2^{n-1}$ middle thirds of length $\frac{1}{3^n}$, and so the total length of $[0,1] \setminus C$ must be

$$\begin{aligned}\frac{1}{3}+2\left(\frac{1}{9}\right)+4\left(\frac{1}{27}\right)+\cdots+2^{n-1}\left(\frac{1}{3^n}\right)+\cdots &= \sum_{n=1}^{\infty}\frac{2^{n-1}}{3^n}=\frac{1}{3}\sum_{n=1}^{\infty}\left(\frac{2}{3}\right)^{n-1}\\ &= \frac{1}{3}\cdot\frac{1}{1-2/3}=1.\end{aligned}$$

Thus the Cantor set has zero length. □

Note that at this point we have not defined what we mean by a "measure zero" set. Thus, we cannot claim that the above argument implies that the Cantor set $C$ has "measure zero." However, if one assumes that "measure" is additive, then $C$ has measure zero because its complement has measure 1. Later, in the section on the Riemann Integral, we define a "measure zero set" and show that the Cantor set is actually a measure zero set.

**Proposition 9.** *The Cantor set $C$ is compact.*

*Proof.* The set $C$ is closed because it is the intersection of closed sets; it is also bounded since it is contained in $[0,1]$. By the Heine–Borel Theorem, the cantor set $C$ is compact. □

To prove the Cantor set is uncountable, we need the following lemma.

**Lemma 7.** *The collection of all sequences of 0's and 1's is uncountable.*

*Proof.* Take a sequence $\{a_n\}$ of 0's and 1's, then $\sum_{n=1}^{\infty}\frac{a_n}{2^n}$ represents an element in $[0,1]$, and conversely each element of $[0,1]$ can be so represented. That is, the map

$$a_n \mapsto 0.a_1a_2a_3\ldots \quad \text{(base 2) is onto.}$$

Hence the set of all $0-1$ sequences, written as $\{0,1\}^{\mathbb{N}}$, has cardinality at least that of the interval $[0,1]$. □

**Proposition 10.** *The Cantor set $C$ is uncountable.*

*Proof.* There are several proofs that $C$ is uncountable, for example, in [20], pp. 309–313, and using the Schröder–Bernstein Theorem, it is proved that $C$ and $[0,1]$ have the same cardinality. Here we start by claiming that every $c \in C$ yields a sequence $(x_1, x_2, x_3, \ldots)$ of zeros and ones, likewise every such sequence corresponds to a point in $C$. By the above lemma since the set of sequences of zeros and ones is uncountable, $C$ must be uncountable. To see intuitively that there is a one-to-one correspondence between $C$ and the sequence $\{x_n\}$, where $x_n = 0$ or 1, first label the intervals $C_1, C_2$, and so on as left (L) and right (R) as shown in Figure 1.25.

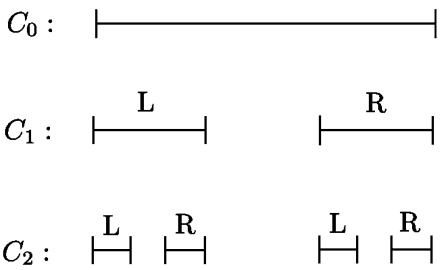



Figure 1.25: Labelling of the intervals as L or R

For each $c \in C$, set

$$x_1 = 0 \text{ if } c \text{ is in the (L) component of } C_1.$$

$$x_1 = 1 \text{ if } c \text{ is in the (R) component of } C_1.$$

Having established where in $C_1$ the point $c$ is located, now we have two possible components of $C_2$ that might contain $c$. Next we set

$$x_2 = 0 \text{ if } c \text{ is in the (L) component of } C_2.$$

$$x_2 = 1 \text{ if } c \text{ is in the (R) component of } C_2.$$

Continuing in this fashion, for every $c \in C$ we can generate a sequence $\{x_1, x_2, \ldots\}$ of 0's or 1's, a set we already know to be uncountable. □

Thus we can write

$$card(C) = card(2^{\mathbb{N}}) = card([0,1]).$$

Here by "$card(A)$" we mean the cardinality of the set $A$. This is hard to believe that the Cantor set is as big as $[0,1]$ but has length 0. No wonder the Cantor set sometimes referred to as "Cantor dust."

To motivate Theorem 36, we consider $x \in [0,1]$, where $x$ can be written as $x = 0.a_1a_2a_3\ldots$ (base 3) with each $a_n = 0, 1$, or 2. These three choices correspond to a three-way splitting of intervals. For example, the three intervals for $C_1$ are $[0, \frac{1}{3}]$, $(\frac{1}{3}, \frac{2}{3})$, and $[\frac{2}{3}, 1]$ (Figure 1.26).

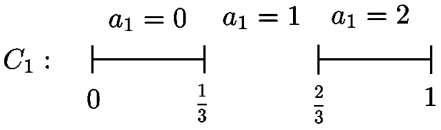



Figure 1.26: Splitting of intervals in $C_1$

We need to pay special attention to the end points, since $1/3 = 0.1 = 0.0222\ldots$ (base 3), $2/3 = 0.2 = 0.1222\ldots$ (base 3), and $1 = 1.0 = 0.222\ldots$

(base 3), but each has at least one representation with $a_1$ in the proper range. Next we examine $C_2$ but ignore the discarded intervals (Figure 1.27):

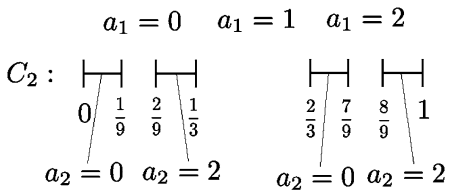



Figure 1.27: Splitting of intervals in $C_2$

Again there is some ambiguity at the end points, for example,

$$7/9 = 0.21000\cdots = 0.20222\ldots$$

These end points of removed intervals are called triadic rational numbers of the form $\frac{m}{3^n}$. Even though these points have two distinct ternary expansions, exactly one of the expressions has all digits $a_k = 0$ or 2. These examples point us to the following which we state without the proof (see [14], p. 28).

**Theorem 36.** *$c \in C$ if and only if $c$ can be written as*

$$c = 0.a_1a_2a_3\cdots = \sum_{n=1}^{\infty} \frac{a_n}{3^n},$$

*where each $a_n$ is either 0 or 2.*

**Remark 21** (Dimension of $C$)**.** The Cantor set is as "big" as the interval $[0, 1]$ or the set of reals $\mathbb{R}$ in the sense that they are all uncountable, but it is also as small as a point because a single point and the Cantor set have zero length. Does it make sense to ask about the dimension of $C$? Even though we did not give a proper definition of the dimension, we can have a discussion toward understanding its dimension. Actually it turns out that the dimension of $C$ is $\frac{\ln 2}{\ln 3}$, a non-integer, or fractional dimension. The notion of a non-integer or fractional dimension is the motivation behind the term "fractal." The Cantor set $C$ is a fractal. To calculate the dimension of $C$, consider the following sets and their dimensions denoted $d$ in Figure 1.28.

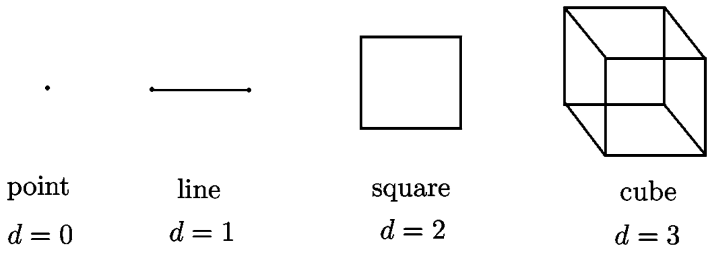



Figure 1.28: A point, line segment, square, and cube

Now think about what happens when you magnify a point by 3; nothing changes, you still have one single point or $3^0 = 1$ copy of itself. However a segment magnified by 3 results in $3^1 = 3$ copies of itself and a square, which is two-dimensional, magnified three times results in $3^2 = 9$ copies. Now let us consider the Cantor set $C$. We start with $[0, 1]$; if we magnify by 3, we get $[0, 3]$, and if you delete the middle third $(1, 2)$, then you end up with $[0, 1] \cup [2, 3]$. Thus you get two of the original copies (Figure 1.29).

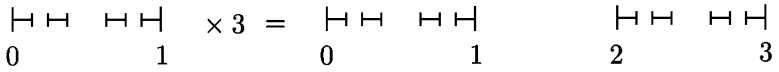



Figure 1.29: Magnifying the Cantor set by a factor of 3

Let $d$ denote the dimension of the Cantor set; then, solving $2 = 3^d$ for $d$ yields $d = \dfrac{\ln 2}{\ln 3} \approx .631$.

Note also that there are other interesting objects whose construction is reminiscent of the Cantor set. For example, Sierpinski's triangle is obtained by starting with a closed (filled) unit equilateral triangle and in the first step removing one open triangle of size $1/2$, in the second step removing 3 open triangles of size $1/4$, and thus at the $n$th step removing $3^{n-1}$ open triangles of size $1/2^n$. Sierpinski's triangle also has a fractional dimension $d = \dfrac{\ln 3^n}{\ln 2^n} = \dfrac{\ln 3}{\ln 2} = 1.585$ (Figure 1.30).

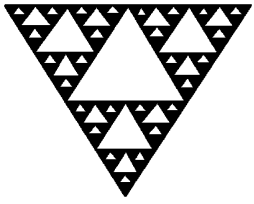


Figure 1.30: Sierpinski's triangle

## Exercises

1. Show that $\emptyset$ and $\mathbb{R}$ are the only subsets of $\mathbb{R}$ that are both open and closed in $\mathbb{R}$.

2. Discuss whether the following sets are open or closed. Determine the interiors, closures, and the boundaries of each set.

   a) $(1, 4)$ in $\mathbb{R}$.

   b) $[2, 5]$ in $\mathbb{R}$.

   c) $\{r \in (0, 1) : \ r \text{ is rational}\}$ in $\mathbb{R}$.

   d) $\displaystyle\bigcap_{n=1}^{\infty} \left[-1, \frac{1}{n}\right)$ in $\mathbb{R}$.

3. Find the accumulation points of the following sets in $\mathbb{R}$:

   a) $A = \mathbb{Q}$.

   b) $A = \mathbb{Z}$.

   c) $A = (0, 1)$.

   d) $A = \left\{(-1)^n + \frac{1}{n} : \ n \in \mathbb{N}\right\}$.

4. Identify which of the following sets are compact. Which are connected?

   a) A finite set in $\mathbb{R}$

   b) $\left\{\frac{1}{n} : \ n \in \mathbb{N}\right\} \cup \{0\}$

   c) $\{x \in \mathbb{R} : \ 0 \leq x \leq 1 \text{ and is irrational}\}$

   d) A closed set in $[0, 1]$

   e) The boundary of a bounded set in $\mathbb{R}$

   f) $\mathbb{Z}$ in $\mathbb{R}$

5. Let $A$ and $B$ be compact subsets of $\mathbb{R}$. Prove that $A \cup B$ and $A \cap B$ are compact.

6. Suppose $K \subset \mathbb{R}$ is compact and nonempty. Prove that $\sup K, \ \inf K \in K$.

7. Prove that the intersection of connected sets in $\mathbb{R}$ is connected. Show that this is false if $\mathbb{R}$ is replaced by $\mathbb{R}^2$.

8. Let $K$ be a nonempty compact set. Let $\{A_n\}$ be a nonempty decreasing sequence of closed subsets of $K$. Prove that $\bigcap_{n\geq 1} A_n$ is not empty (Cantor's Intersection Theorem).

9. Let $C$ be the Cantor set as defined above. Prove that $C$ is totally disconnected, that is, if $x, y \in C$ and $x \neq y$, then $x \in U$ and $y \in V$, where $U$ and $V$ are open sets that disconnect $C$.

10. Prove that if $A \subset \mathbb{R}$ is connected and $A \subset B \subset \overline{A}$, then $B$ is connected.

11. Let $A$ be a subset of $\mathbb{R}$ and $B$ be a set of points $x \in \mathbb{R}$ with the property that $A \cap (x - \delta, x + \delta)$ is uncountable for every $\delta > 0$. Show that $A \setminus B$ is finite or countable.

12. Let $A \subset \mathbb{R}$ be uncountable. Show that $A$ has at least one accumulation point.

## 1.6 Continuous Functions

### Functional Limits

Let $f : S \subseteq \mathbb{R} \to \mathbb{R}$, and recall that a limit point $a$ of $S$ is a point where every $\epsilon$-neighborhood $V_\epsilon(a)$ of $a$ intersects $S$ in some point other than $a$. We showed in the section on the topology of $\mathbb{R}$ that this is equivalent to the claim that $a$ is a limit point of $S$ if and only if $a = \lim x_n$ for some sequence $\{x_n\}$ in $S$ with $x_n \neq a$. Furthermore, limit points of $S$ do not necessarily belong to $S$ unless $S$ is closed. Let $a$ be a limit point of $S$; then we write

$$\lim_{x \to a} f(x) = L,$$

by which we roughly mean that values of $f(x)$ get arbitrarily close to the real number $L$ when $x$ is approaching $a$. Moreover the point $a$ need not even be in the domain of $f$; thus we do not even think of the case when $x = a$. The following definition is very "similar" to the definition for the limit of a sequence.

**Definition 29** ($\epsilon - \delta$ Definition of Continuity)**.** Let $f : S \subseteq \mathbb{R} \to \mathbb{R}$ and let $a$ be a limit point of the domain $S$. We say $\lim_{x \to a} f(x) = L$ if for each $\epsilon > 0$ there exists a number $\delta > 0$ such that

$$|f(x) - L| < \epsilon \quad \text{whenever} \quad 0 < |x - a| < \delta \text{ (and } x \in S).$$

Note that the condition $0 < |x - a|$ simply states $x \neq a$ and the statement

$$|f(x) - L| < \epsilon \quad \text{is equivalent to} \quad L - \epsilon < f(x) < L + \epsilon \quad \text{or} \quad f(x) \in V_\epsilon(L).$$

Similarly, the statement

$$|x - a| < \delta \quad \text{is satisfied if and only if} \quad x \in V_\delta(a).$$

Thus the above definition can be restated in terms of neighborhoods and clarifies the geometry behind the above definition.

**Definition 30** (Topological Definition)**.** Let $f : S \subseteq \mathbb{R} \to \mathbb{R}$ and let $a$ be a limit point of the domain $S$. We say $\lim_{x \to a} f(x) = L$ if for every $\epsilon$-neighborhood $V_\epsilon(L)$ of $L$, there exists a $\delta$-neighborhood $V_\delta(a)$ around $a$ such that for all $x \neq a$ (with $x \in S$) and $x \in V_\delta(a)$, it follows that $f(x) \in V_\epsilon(L)$.

**Example 39.**

a) Let $f : \mathbb{R} \to \mathbb{R}$ be defined as $f(x) = k$ for all $x \in \mathbb{R}$, where $k$ is some constant. Then for any $a \in \mathbb{R}$, $\lim_{x \to a} f(x) = k$. Indeed, given any $\epsilon > 0$, let $\delta = 1$ (or any other positive number). Then

$$|f(x) - k| = |k - k| = 0 < \epsilon \quad \text{whenever} \quad 0 < |x - a| < 1.$$

b) Let $f(x) = 4x + 1$; we claim that $\lim_{x \to 1} f(x) = 5$. Let $\epsilon > 0$ be given; the $\epsilon - \delta$ definition requires finding a $\delta > 0$ so that $|f(x) - 5| < \epsilon$ whenever $0 < |x - 1| < \delta$. Notice that

$$|f(x) - 5| = |(4x + 1) - 5| = |4x - 4| = 4|x - 1|.$$

Therefore, if we choose $\delta = \dfrac{\epsilon}{4}$, then $0 < |x - 1| < \delta$ implies that

$$|f(x) - 5| < 4\left(\frac{\epsilon}{4}\right) = \epsilon.$$

c) Let $f(x) = x^2$; we claim that $\lim_{x \to 2} f(x) = 4$. Given $\epsilon > 0$ as in the previous examples, our aim is to show $|f(x) - 4| < \epsilon$ whenever $0 < |x - 2| < \delta$, and the question is how to choose this $\delta$. Again we start by looking at

$$|f(x) - 4| = |x^2 - 4| = |(x - 2)(x + 2)| = |x - 2||x + 2|.$$

We can make $|x - 2|$ as small as possible, but we need to find an upper bound for the term $|x + 2|$. If we decide to take $\delta = 1$, then the $\delta$-neighborhood of $a = 2$ must have radius at most 1, and this makes it possible for us to say $|x + 2| \leq |3 + 2| = 5$ holds true for all $|x - 2| < \delta$. Now choose $\delta = \min\{1, \frac{\epsilon}{5}\}$, then

$$|x^2 - 4| = |x - 2||x + 2| < 5\left(\frac{\epsilon}{5}\right) = \epsilon.$$

When we wrote about the convergence of sequences in Section 1.4 of Chapter 1, we covered important theorems such as the algebraic limit theorem and order limit theorem. It is natural to ask whether or not analogous statements are valid for functional limits.

**Theorem 37** (Algebraic Limit Theorem for Functional Limits)**.** *If $f$ and $g$ are defined on a common set $S$ and if $\lim_{x\to a} f(x) = L$ and $\lim_{x\to a} g(x) = M$ for some limit point $a$ of $S$, then*

*a)* $\lim_{x\to a} f(x) + g(x) = L + M.$

*b)* $\lim_{x\to a} f(x)\, g(x) = LM.$

*c)* $\lim_{x\to a} \dfrac{f(x)}{g(x)} = \dfrac{L}{M}$ *provided* $M \neq 0.$

*d)* $\lim_{x\to a} kf(x) = kL$ *for all* $k \in \mathbb{R}.$

The proof follows from the algebraic limit theorem for sequences.

**Theorem 38** (Sequential Criterion for Functional Limits)**.** *Let $f : S \to \mathbb{R}$ and let $a$ be an accumulation point of $S$. The following statements are equivalent:*

*a)* $\lim_{x\to a} f(x) = L.$

*b) For all sequences $\{x_n\} \subseteq S$ satisfying $x_n \neq a$ and $\{x_n\}$ converging to $a$, it follows that the sequence $\{f(x_n)\}$ converges to $L$.*

*Proof.* Assume part a) holds. Choose a sequence $\{x_n\}$ satisfying $x_n \neq a$ and $\lim_{n\to\infty} x_n = a$. Let $\epsilon > 0$ be given. Then there exists $\delta > 0$ such that $|f(x) - L| < \epsilon$ if $x \in S$ and $0 < |x - a| < \delta$. Since we are assuming that $\{x_n\}$ is converging to $a$, there exists $N$ such that $n \geq N$ implies that $0 < |x_n - a| < \delta$. Thus, for those $n \geq N$, we have $|f(x_n) - L| < \epsilon$, which proves part b).

The reverse implication is done by contradiction. Suppose part a) is false, i.e., suppose $L$ is not a limit of $f$ at $a$. We must find a sequence $\{x_n\}$ in $S$ that converges to $a$ with each $x_n \neq a$, such that $\{f(x_n)\}$ does not converge to $L$. Since $L$ is not a limit of $f$ at $a$, there exists an $\epsilon > 0$ such that for every $\delta > 0$ there exists an $x \in S$ with $0 < |x - a| < \delta$ such that $|f(x) - L| \geq \epsilon$. In particular, for each $n \in \mathbb{N}$, there exists $x_n \in S$ with

$$0 < |x_n - a| < \frac{1}{n} \quad \text{such that} \quad |f(x_n) - L| \geq \epsilon.$$

Now notice that the sequence $\{x_n\}$ converges to $a$ with $x_n \neq a$ for all $n$, but $\{f(x_n)\}$ cannot converge to $L$. □

**Corollary 5.** *If $f$ has limit at $a$, this limit is unique.*

This follows from sequential criterion of functional limits and the uniqueness of the limit for sequences (see Theorem 38). One very useful application of the sequential criterion for the limit of a function is to show that a given limit does not exist. This is illustrated in the following example.

**Example 40.** Consider the function $f(x) = \sin\left(\frac{1}{x}\right)$ for $x > 0$. We claim that $\lim_{x\to 0} f(x)$ does not exist. Let $x_n = \frac{2}{n\pi}$ for all $n \in \mathbb{N}$, then $\lim_{n\to\infty} x_n = 0$. However,

$$\sin\left(\frac{n\pi}{2} + 2\pi k\right) = \begin{cases} 0 & \text{if} \quad n = 0 \\ 1 & \text{if} \quad n = 1 \\ 0 & \text{if} \quad n = 2 \\ -1 & \text{if} \quad n = 3. \end{cases}$$

Clearly $\{f(x_n)\}$ is the sequence $\{1, 0, -1, 0, 1, 0, -1, \dots\}$, which does not converge.

## Continuity

The word *continuous* in everyday use means "unbroken," "without a break," or "without gaps." In mathematics it describes functions for which small changes in the input result in small changes in the output. But these are vague descriptions; continuity needs to be expressed in terms of rigorous mathematical language. The following definition does exactly this.

**Definition 31.** A real-valued function $f : S \subseteq \mathbb{R} \to \mathbb{R}$ is said to be continuous at the point $a \in S$ if for each $\epsilon > 0$ there exists a number $\delta > 0$ such that

$$|f(x) - f(a)| < \epsilon \quad \text{for all} \quad x \in S \quad \text{with} \quad |x - a| < \delta.$$

Written in slightly different terms, this definition requires

$$f(V_\delta(a)) \subset V_\epsilon(f(a)).$$

That is, $f$ maps a sufficiently small neighborhood of $a$ into some neighborhood of $f(a)$. If $f$ is continuous at every point in the domain $S$, then we say $f$ is *continuous on* $S$ (Figure 1.31).

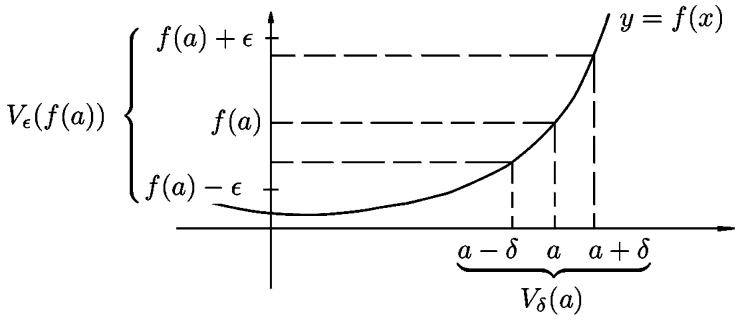



Figure 1.31: Continuity of $f$ at $a$

The definition of continuity resembles the definition of functional limits; however, functional limits require the point $a$ to be a limit point of $S$, and this is not assumed here. In the above definition it is required that the point $a$ be in the domain of $f$; the value $f(a)$ is then the value of $\lim_{x \to a} f(x)$. Note also that the above definition implies any function is continuous at isolated points in its domain. Recall that a point of $S$ that is not a limit point is called an *isolated* point if there is $\delta > 0$ such that if $|x - a| < \delta$ and $x \in S$, then $x = a$. Thus whenever

$$|x - a| < \delta \text{ and } x \in S, \text{ we have } |f(x) - f(a)| = 0 < \epsilon.$$

**Example 41.** Let $f : \mathbb{R} \to \mathbb{R}$ be the identity function $f(x) = x$. Fix a point $x_0 \in \mathbb{R}$. Given $\epsilon > 0$, we must find $\delta > 0$ such that $|f(x) - f(x_0)| < \epsilon$ whenever $|x - x_0| < \delta$. Choosing $\epsilon = \delta$ does the job.

**Example 42.** Let

$$f(x) = \begin{cases} x \sin\left(\dfrac{1}{x}\right) & : x \neq 0 \\ 0 & : x = 0. \end{cases}$$

We claim that $f$ is a continuous function at 0. Observe that

$$|f(x) - f(0)| = |x \sin\left(\frac{1}{x}\right)| \leq |x| \quad \text{for all } x,$$

and given $\epsilon > 0$, we may set $\delta = \epsilon$. Then when $|x - 0| < \delta$, we have

$$|f(x) - f(0)| \leq |x| < \delta = \epsilon.$$

Hence $f$ is continuous at 0. The graph of this function looks like (Figure 1.32):

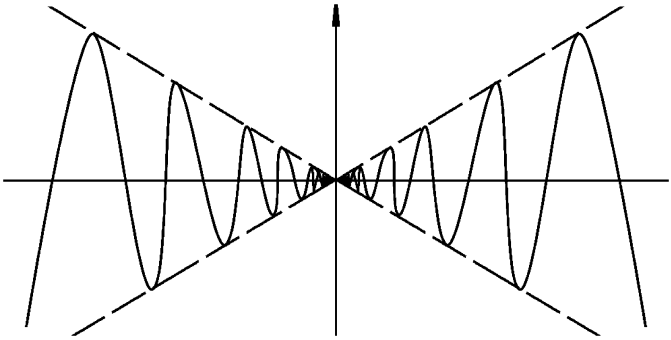


Figure 1.32: $x \sin\left(\frac{1}{x}\right)$ when $x \neq 0$ and $f(0) = 0$

Actually this function is continuous on all points of $\mathbb{R}$.

**Example 43.** In discussing continuity one has to pay attention to the domain of the function. For example, the function $f : \mathbb{R} \to \mathbb{R}$ defined as

$$f(x) = \begin{cases} 0 & : x \notin \mathbb{Q} \\ 1 & : x \in \mathbb{Q} \end{cases}$$

is not continuous at any point of $\mathbb{R}$. Note that for an arbitrary $a \in \mathbb{R}$ every neighborhood of $a$ contains rational points at which $f(x) = 1$ and also irrational points at which $f(x) = 0$. Thus $\lim_{x \to a} f(x)$ cannot possibly exist. Thus $f$ is discontinuous at every $a \in \mathbb{R}$. However, if we restrict $f$ to be a function from $\mathbb{Q}$ to $\mathbb{Q}$, then $f(x) = 1$ is a constant function and is continuous at every point of $\mathbb{Q}$. It is quite difficult to graph (see our attempt below in Figure 1.33). This function is sometimes referred to as Dirichlet's function.

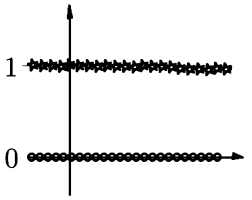



Figure 1.33: Dirichlet's function

## Properties of Continuous Functions

It is easy to prove that if $f$ and $g$ are defined on a common set $S$ and are continuous at a point $a \in S$, then their sum $f+g$ and product $fg$ are continuous at $a$, and the quotient $\frac{f}{g}$ is continuous whenever $g(a) \neq 0$. Since $f(x) = x$ is continuous, this allows us to conclude that every polynomial $p(x) = c_0 + c_1 x +$

$c_2x^2 + \cdots + c_nx^n$ is continuous on $\mathbb{R}$. Rational functions, being the quotient of two polynomials, are continuous except at the points where the denominator is zero. Furthermore, given two functions

$$f : S_1 \to \mathbb{R} \quad \text{and} \quad g : S_2 \to \mathbb{R},$$

assume that $f(S_1) \subset S_2$, where $f(S_1) = \{f(x) : \ x \in S_1\}$, so that we can define the *composition* $(g \circ f)(x) = g(f(x))$ on $S_1$. If $f$ is continuous at $a \in S_1$ and $g$ is continuous at $f(a) \in S_2$, then $g \circ f$ is continuous at $a$. Proofs of these facts are left to the reader as exercises. Because of these facts we can safely state that the function $h(x) = \sqrt{2x^2 + 1}$ is continuous. Notice that $h(x)$ is the composition of two continuous functions, namely $g(x) = \sqrt{x}$ and the polynomial function $f(x) = 2x^2 + 1$.

The $\epsilon - \delta$ characterization of continuity given in the above definition is not the only one; there are several equivalent ways to characterize continuity.

**Theorem 39** (Characterization of Continuity)**.** *Let $f : S \subseteq \mathbb{R} \to \mathbb{R}$ and let $a \in S$. Then the following three conditions are equivalent:*

*a) $f$ is continuous at $a$.*

*b) If $\{x_n\}$ is any sequence in $S$ such that $\{x_n\}$ converges to $a$, then*

$$\lim_{n\to\infty} f(x_n) = f(a).$$

*c) For every neighborhood $V_\epsilon(f(a))$, there exists a $V_\delta(a)$ such that*

$$f(S \cap V_\delta(a)) \subseteq V_\epsilon(f(a)).$$

*Furthermore, if $a$ is a limit point of $S$, then the above are all equivalent to following:*

*d) $f$ has a limit at $a$ and $\lim_{x\to a} f(x) = f(a)$.*

*Proof.* The statement a) is the $\epsilon - \delta$ definition of continuity. Suppose $a$ is an isolated point of $S$. Then there exists a neighborhood $V_\delta(a)$ of $a$ such that $V_\delta(a) \cap S = \{a\}$. It follows that, for any neighborhood $V_\epsilon(f(a))$ of $f(a)$,

$$f(S \cap V_\delta(a)) \subseteq \{f(a)\} \subseteq V_\epsilon(f(a)).$$

Thus c) always holds. Similarly it is not difficult to see that if $\{x_n\}$ is a sequence in $S$ converging to $a$, then $x_n \in V_\delta(a)$ for all $n \geq N$, which implies that $x_n = a$ for all $n \geq N$, so $\lim_{n\to\infty} f(x_n) = f(a)$. Thus b) holds, and a), b), and c) are all equivalent. Now, let us look at the case when $a$ is a limit point of $S$. Then a) $\Leftrightarrow$ d) is Definition 29, d) $\Leftrightarrow$ c) is Definition 30, and d) $\Leftrightarrow$ b) is very much like Theorem 38. □

Statement b) in the above theorem is called the *sequential characterization* of continuity and is also a useful characterization of *discontinuity*. Given in the following Corollary 6.

**Corollary 6.** *Let $f : S \subseteq \mathbb{R} \to \mathbb{R}$ and let $a \in S$. Then $f$ is discontinuous at $a$ if and only if there exists a sequence $\{x_n\}$ in $S$ such that $\{x_n\}$ converges to $a$, but the sequence $\{f(x_n)\}$ does not converge to $f(a)$.*

**Example 44.** Let

$$f(x) = \begin{cases} \sin\left(\dfrac{1}{x}\right) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

Let $\{x_n\} = \dfrac{1}{(n+1/2)\pi}$, then for $x \neq 0$, $\{f(x_n)\} = \{\sin(n+1/2)\pi\} = (-1)^n$, which is an oscillating sequence. Its limit does not exist and so $f$ is a discontinuous at 0.

Recall that inverse image (or pre-image) of a set $A \subset \mathbb{R}$ under the function $f : S \to \mathbb{R}$ is defined to be the set $\{x \in S : \ f(x) \in A\}$, and usually it is written as $f^{-1}(A)$. The inverse image of a set under any mapping always makes sense. Although the notation is similar, inverse functions have nothing to do with inverse images. The inverse of a function might not exist. We have seen that the $\epsilon - \delta$ definition of continuity can be expressed using neighborhoods as $f(V_\delta(a)) \subset V_\epsilon(f(a))$. Stated in terms of inverse image, our condition reads (Figure 1.34)

$$V_\delta(a) \subset f^{-1}(V_\epsilon(f(a)).$$

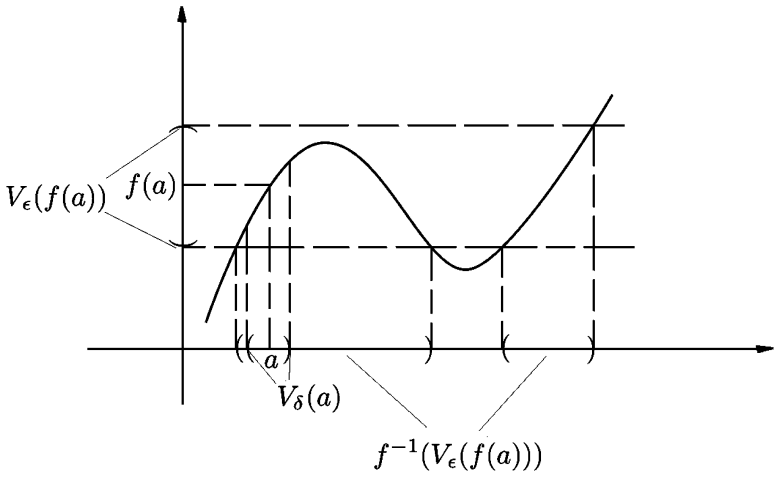



Figure 1.34: Continuity and the inverse image of a set

The following theorem illustrates how continuity is related to the pre-image of open subsets in the range of the function. This characterization of continuity is useful not only for real-valued functions but in a more general setting as well. (See [26], p. 180.)

**Theorem 40.** *A function $f : S \subseteq \mathbb{R} \to \mathbb{R}$ is continuous on $S$ if and only if for any open set $O \subset \mathbb{R}$ there exists an open set $G \subset \mathbb{R}$ such that $f^{-1}(O) = G \cap S$.*

In case where the domain of $f$ is all of $\mathbb{R}$, the previous theorem may be restated as follows:

**Corollary 7.** *A function $f : \mathbb{R} \to \mathbb{R}$ is continuous if and only if $f^{-1}(O)$ is open in $\mathbb{R}$ whenever $O$ is open in $\mathbb{R}$.*

For the proof of the above theorem and its corollary, we refer the reader to [26], p. 180. As shown in the following figure, if $f$ is not continuous, then the inverse image of an open set is not necessarily open (Figure 1.35).

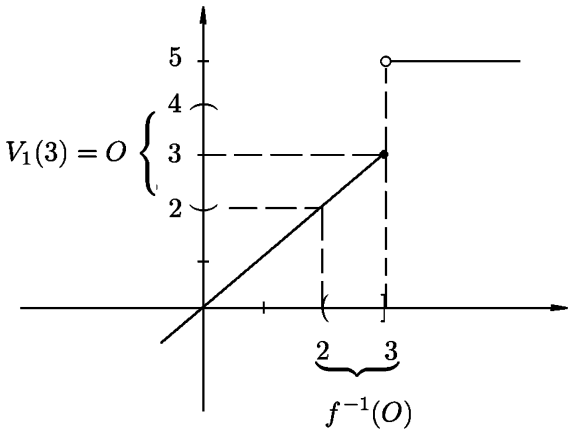



Figure 1.35: The set O is open, but $f^{-1}(O)$ is not open

Note that if $A \subseteq \mathbb{R}$ is an open subset in the domain of $f$, even if $f$ is continuous, we cannot claim that image $f(A)$ is open. A typical counterexample is the continuous function $f(x) = x^2$. If we set $A = (-1, 1)$, then $f(A) = [0, 1)$, which is not open.

## Continuous Functions and Compact Sets

**Definition 32.** A function $f : S \to \mathbb{R}$ is said to be *bounded* if its range $f(S)$ is a bounded subset of $\mathbb{R}$. That is, $f$ is bounded if there exists a real number $M$ such that

$$|f(x)| \leq M \quad \text{for all} \quad x \in S.$$

Unfortunately, a continuous function could be unbounded even if its domain is bounded. For example, set $S = (0, 1)$ and $f(x) = \dfrac{1}{x}$, and then $f$ is continuous in its domain, but its range $f(S) = (1, \infty)$ is an unbounded set. If it happens that the domain of a continuous function is both "closed and bounded"—that is to say a "compact" set (recall from the Heine–Borel Theorem that a subset

of $\mathbb{R}$ is compact if and only if it is closed and bounded)—then the function is bounded.

**Theorem 41** (Preservation of Compact Sets)**.** *Let $K$ be a compact subset of $\mathbb{R}$, and let $f : K \to \mathbb{R}$ be continuous. Then $f(K)$ is compact as well.*

*Proof.* Let $\{U_\alpha\}$ be an open cover for $f(K)$. Since $f$ is continuous, the inverse image of an open set is open; thus each of the sets $f^{-1}(U_\alpha)$ is open. However $K$ is compact, and therefore every open cover has a finite sub cover, i.e., there are finitely many indices $\alpha_1, \alpha_2, \ldots, \alpha_n$, such that

$$K \subset f^{-1}(U_{\alpha_1}) \cup f^{-1}(U_{\alpha_2}) \cup \cdots \cup f^{-1}(U_{\alpha_n}). \tag{1.2}$$

Recalling the fact $f(f^{-1}(E)) \subset E$ for every subset $E$ in $\mathbb{R}$, using the inclusion in (1.2), we get

$$f(K) \subset U_{\alpha_1} \cup U_{\alpha_2} \cup \cdots \cup U_{\alpha_n}. \qquad \square$$

Note that the inverse image of a compact set under a continuous function need not be compact. For example, if we set $E = \{0\} \subset \mathbb{R}$ and $f : \mathbb{R} \to \mathbb{R}$ to be the function $f(x) = 0$ for all $x \in \mathbb{R}$, then $f^{-1}(E) = \mathbb{R}$, which is not compact. The above theorem has important consequences. For example, since compact subsets of $\mathbb{R}$ are closed and bounded (Heine–Borel Theorem), if $K$ is a compact subset of $\mathbb{R}$ and $f : K \to \mathbb{R}$ is continuous, then $f(K)$ is closed and bounded. Thus $f$ is bounded. For the other important consequence of Theorem 41, we need the following definition.

**Definition 33.** We say that $f : S \subseteq \mathbb{R} \to \mathbb{R}$ has a *maximum* at the point $x_0 \in S$ if $f(x) \leq f(x_0)$ for all $x \in S$. Similarly, $f$ has a *minimum* at the point $x_1 \in S$ if $f(x_1) \leq f(x)$ for all $x \in S$.

**Corollary 8** (Extreme Value Theorem)**.** *If $f$ is continuous on a closed and bounded set $K \subseteq \mathbb{R}$, then it attains its maximum and minimum on $K$; in other words, there are $x_0, x_1 \in K$ such that $f(x_1) \leq f(x) \leq f(x_0)$ for all $x \in K$.*

*Proof.* Let $K$ be a compact subset of $\mathbb{R}$, and suppose $f : K \to \mathbb{R}$ is continuous. From Theorem 41 it follows that $f(K)$ is compact. Thus $f(K)$ has both maximum and minimum ($\inf K = \min K$ and $\max K = \sup K$ ). Let $y_0 = \max f(K)$ and $y_1 = \min f(K)$, then there exist $x_0, x_1 \in K$ such that $y_0 = f(x_0)$ and $y_1 = f(x_1)$. It follows that $f(x_1) \leq f(x) \leq f(x_0)$ for all $x \in K$. $\square$

**Remark 22.** To apply the Extreme Value Theorem, it is essential for $K$ to be both closed and bounded. If $K$ is not closed, say $K = (0, 1)$, we can find a continuous function $f : (0, 1) \to \mathbb{R}$ such as $f(x) = x$, where $f$ does not assume its max or min values. If $K$ is not bounded, say $K = [0, \infty)$, then the identity function $f : [0, \infty) \to \mathbb{R}$ is continuous but does not assume a maximum value. An example of unbounded continuous function $f(x) = \dfrac{1}{x}$ is given in Figure 1.36,

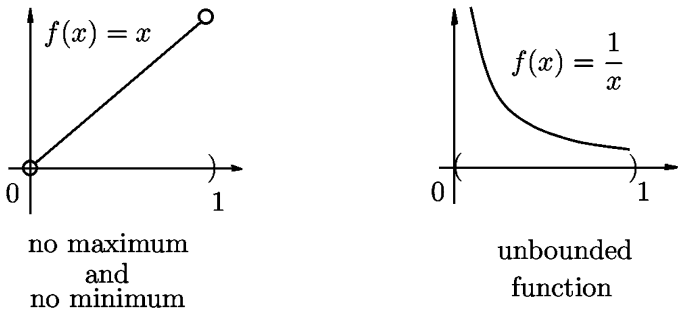



Figure 1.36: Two continuous functions on a noncompact set

**Remark 23.** Recall that when we say "$f$ is continuous on $S$," we mean it is continuous at each point $a$ of $S$. In other words, given $\epsilon > 0$, we can find $\delta > 0$ such that $|f(x) - f(a)| < \epsilon$ whenever $|x - a| < \delta$. Here $\delta$ may depend on the point $a$ as well as on $\epsilon$, in general $\delta = \delta(\epsilon, a)$. For example, consider $f(x) = x^2$ and take an arbitrary $a \in \mathbb{R}$, then $|f(x) - f(a)| = |x^2 - a^2| = |x - a||x + a|$, and it turns out $\delta = \dfrac{\epsilon}{2|a| + 1}$ does the trick (why? Fill in the details) (Figure 1.37).

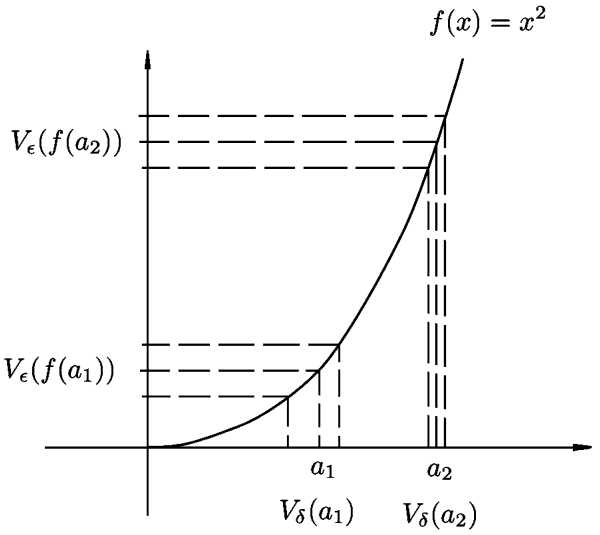



Figure 1.37: Larger $a$ requires smaller $\delta$

In the above figure for $f(x) = x^2$, we see that a larger $a$ requires smaller $\delta$, thus $\delta = \delta(a, \epsilon)$. There is a stronger concept of continuity called *uniform*

*continuity*, basically requiring that given $\epsilon > 0$, a single $\delta > 0$ can be chosen that works simultaneously for all the points $a \in S$, i.e., $\delta = \delta(\epsilon)$ only. A function $f$ defined on a set $S \subset \mathbb{R}$ is said to be *uniformly continuous* on $S$, if for each $\epsilon > 0$ there is a $\delta > 0$ such that $|f(x) - f(a)| < \epsilon$ for all pairs of point $x, a \in S$ with $|x - a| < \delta$. Every uniformly continuous function is continuous (pointwise), but the converse is not true. For example, while $f(x) = 3x$ is uniformly continuous on $\mathbb{R}$, $f(x) = x^2$ is continuous but not uniformly continuous on $\mathbb{R}$. However, if the domain of the function is a compact set, then we have the following theorem:

**Theorem 42.** *Suppose $f : K \to \mathbb{R}$ is continuous on a compact set $K$; then $f$ is uniformly continuous on $K$.*

For the proof of this theorem, we refer the reader to, e.g., [26], p. 194. Going back to the above example of the continuous function $f(x) = x^2$, if we restrict the domain of this function to some compact set, say $K = [-3, 3]$, then $|x + a| \leq 6$, thus

$$|f(x) - f(a)| = |x^2 - a^2| = |x - a||x + a| \leq 6\delta = \epsilon.$$

Thus, $f(x) = x^2$ is uniformly continuous on $[-3, 3]$.

## Continuous Functions and Connected Sets

Now we turn our attention to connected subsets of $\mathbb{R}$. The following theorem states that the continuous image of a connected set is connected.

**Theorem 43** (Preservation of Connected Sets)**.** *Let $A$ be a connected subset of $\mathbb{R}$ and $f : A \to \mathbb{R}$ be a continuous function. Then $f(A)$ is a connected subset of $\mathbb{R}$.*

*Proof.* The idea of the proof is that if we let $U$ and $V$ be open sets in $\mathbb{R}$ and assume that $U$ and $V$ disconnect $f(A)$, then it is not difficult to show that the sets $f^{-1}(U)$ and $f^{-1}(V)$ are open sets (since $f$ is continuous) in $\mathbb{R}$ that disconnect $A$. Suppose $f(A)$ is not connected. By definition, we can write $f(A) \subset U \cup V$, where $f(A) \cap U \neq \emptyset$, $f(A) \cap V \neq \emptyset$, and $f(A) \cap U \cap V = \emptyset$. Now, $f^{-1}(U) = U^* \cap A$ for some open set $U^*$, and similarly, $f^{-1}(V) = V^* \cap A$ for some open set $V^*$. From the conditions on $U$ and $V$, we obtain that $U^* \cap V^* \cap A = \emptyset$, $A \subset U^* \cup V^*$, $U^* \cap A \neq \emptyset$, and $V^* \cap A \neq \emptyset$. Thus $A$ is not connected, in contradiction to the hypothesis. □

One consequence of this theorem is the fact that continuous real functions assume all intermediate values on an interval (Figure 1.38).

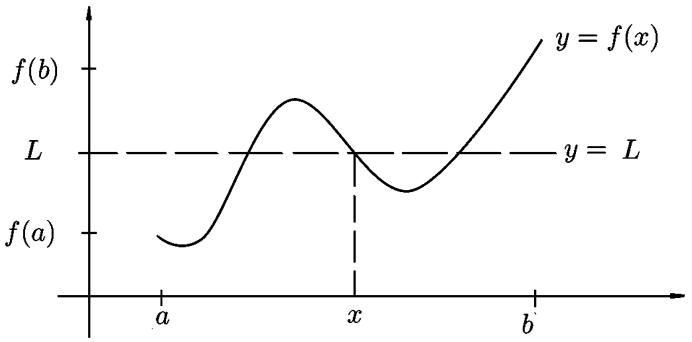



Figure 1.38: The Intermediate Value Theorem

**Corollary 9** (Intermediate Value Theorem)**.** *Let $a < b$ and $f : [a, b] \to \mathbb{R}$ be a continuous function. If $f(a) < f(b)$ and if $L$ is a number such that $f(a) < L < f(b)$, then there exists a point $x \in (a, b)$ such that $f(x) = L$. A similar result holds if $f(a) > f(b)$.*

*Proof.* Since $[a, b]$ is connected, by Theorem 43 we know $f([a, b])$ is a connected subset of $\mathbb{R}$. However, the connected subsets of the real line have a particularly simple structure; they are intervals. A subset $A$ of the real line $\mathbb{R}$ is connected if and only if it has the following property: If $x \in A$, $y \in A$ and $x < k < y$, then $k \in A$. Apply this property to the set $f([a, b])$. □

The Intermediate Value Theorem could be very useful in deciding about the roots of real-valued continuous functions:

**Example 45.**

a) The cubic function $f(x) = x^3 - x - 1$ has a real root. Since $f(1) = -1$ and $f(2) = 5$ and $-1 < 0 < 5$, there exists a $c$ such that $f(c) = 0$.

b) Let $f : [1, 3] \to [0, 6]$ be a continuous function satisfying $f(1) = 0$ and $f(3) = 6$. Then there is a point $c$ in $[1, 3]$ such that $f(c) = c$ (fixed point). To see this define the function $g(x) = f(x) - x$, and since the difference of two continuous function is continuous, $g$ is a continuous function. Moreover, $g(1) = f(1) - 1 = -1$ and $g(3) = f(3) - 3 = 3$. Hence, by the Intermediate Value Theorem, $g$ must vanish at some $c \in [1, 3]$. Thus, $g(c) = f(c) - c = 0$, and this $c$ is a fixed point of $f$.

## Exercises

1. Use the definition of continuity at a point to prove that:

(a) $f(x) = 3x - 5$ is continuous at $x = 2$.

(b) $f(x) = x^2$ is continuous at $x = 3$.

(c) $f(x) = 1/x$ is continuous at $x = 1/2$.

2. A function $f : \mathbb{R} \to \mathbb{R}$ is called *Lipschitz* with Lipschitz constant $\alpha > 0$ if

$$|f(x) - f(y)| \leq \alpha|x - y| \quad \text{for all } x, y \in \mathbb{R}.$$

Give two examples of Lipschitz functions. Moreover, prove that every Lipschitz function is continuous.

3. Let $f : [a, b] \to \mathbb{R}$ be continuous. Prove that $|f|$ is also continuous on $[a, b]$. Is the converse true? Namely, if $|f|$ is continuous on $[a, b]$, is $f$ also continuous on $[a, b]$?

4. Let $f(x) = [x]$ be the greatest integer function that is less than or equal to $x$, and let $g(x) = x - [x]$. Sketch the graphs of $f$ and $g$. Determine the points at which $f$ and $g$ are continuous.

5. Let $I$ be a closed and bounded subset of $\mathbb{R}$ and $f : I \to \mathbb{R}$ be continuous. Prove that $f$ assumes maximum and minimum values on $I$, i.e., there exist $x_1, x_2 \in I$ such that

$$f(x_1) \leq f(x) \leq f(x_2)$$

for all $x \in I$.

6. Prove that if $f : \mathbb{R} \to \mathbb{R}$ is continuous and periodic, then it attains its supremum and infimum.

7. Let $f : [a, b] \to \mathbb{R}$ be continuous, and show that $\sup_{x\in[a,b]} |f(x)|$ is finite.

8. If $f : \mathbb{R} \to \mathbb{R}$ is continuous and $K \subset \mathbb{R}$ is connected, is $f^{-1}(K)$ necessarily connected?

9. Let $X$ be a compact metric space and $f : X \to X$ is an isometry, i.e.,

$$d(f(x), f(y)) = d(x, y).$$

Show that $f$ is a bijection.

10. Which of the following functions on $\mathbb{R}$ are uniformly continuous?

a) $f(x) = x\sin x$.

b) $f(x) = \dfrac{1}{x^2 + 1}$.

11. Prove that if $f(x)$ is a cubic polynomial, then $f$ has a real root. That is, there is an $x_0$ such that $f(x_0) = 0$.

12. Show that if $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuous on all of $\mathbb{R}^n$ and $K \subset \mathbb{R}^n$ is bounded, then $f(K)$ is bounded.

13. Suppose $f$ is continuous on $[0,1]$, and set $I_k = \left[\frac{k-1}{2^n}, \frac{k}{2^n}\right]$ for $k = 1, 2, \ldots, 2^n$. Show that given $\epsilon > 0$ there is a natural number $N$ such that $n \geq N$ implies that

$$\sup_{x \in I_k} f(x) - \inf_{x \in I_k} f(x) < \epsilon, \quad k = 1, 2, \ldots, 2^n.$$

14. A sequence of functions $\{f_n\}$ defined on a set $A \subset \mathbb{R}$ is called *equicontinuous* if for every $\epsilon > 0$ there exists a $\delta > 0$ such that $|f_n(x) - f_n(y)| < \epsilon$ for all $n \in \mathbb{N}$ and $|x - y| < \delta$ in $A$. Let $g_n(x) = x^n$ explain why $g_n(x)$ is not equicontinuous on $[0,1]$. Is each $g_n$ uniformly continuous on [0,1]?

## 1.7 Differentiability on $\mathbb{R}$

**Definition 34.** Let $f$ be a function defined on an open set $I$ containing the point $a$. We say $f$ is *differentiable* at $a$ with derivative $f'(a)$ if the limit

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$

exists. We say $f$ is *differentiable* if it is differentiable at each point $a$ of its domain.

Note that an equivalent way to define the derivative of $f$ at $a$ is to claim

$$f'(a) := \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

exists. The assumption that $f$ is defined on an open interval containing $a$ is made so that the above quotient is defined for all $h \neq 0$.

**Example 46.** We now test this definition on some functions.

a) Let $f(x) = c$, where $c$ is constant; let us find $f'(3)$.

$$f'(3) = \lim_{x \to 3} \frac{f(x) - f(3)}{x - 3} = \lim_{x \to 3} \frac{c - c}{x - 3} = 0.$$

b) Let $f(x) = x^2$. To find $f'(3)$, set

$$f'(3) = \lim_{x \to 3} \frac{f(x) - f(3)}{x - 3} = \lim_{x \to 3} \frac{x^2 - 9}{x - 3} = \lim_{x \to 3} (x + 3) = 6.$$

c) Consider the function

$$f(x) = \begin{cases} 1 & : x \in \mathbb{Q} \\ 0 & : x \notin \mathbb{Q}. \end{cases}$$

Since the limit $\lim_{x\to 0} \dfrac{f(x)}{x}$ fails to exist, $f$ is not differentiable at $x = 0$.

d) To check differentiability for the function

$$f(x) = \begin{cases} x^3 & : x \geq 0 \\ 0 & : x < 0 \end{cases}$$

at $x = 0$, we form

$$\frac{f(x)}{x} = \begin{cases} x^2 & : x \geq 0 \\ 0 & : x < 0 \end{cases}$$

and so $f'(0) = \lim_{x\to 0} \dfrac{f(x)}{x} = 0$.

The derivative is defined in terms of a limit of a particular quotient, but there are several other views of the derivative. For example, the derivative of $f$ at $x = a$ is referred to as the *instantaneous rate of change* of $f$ when $x = a$. In elementary calculus $f'(a)$ is sometimes described as the *slope* of the tangent line to the graph of $f$ at $a$. Clearly, the quotient $\dfrac{f(x) - f(a)}{x - a}$ is the slope of a certain *secant line*, which passes through the points $(x, f(x))$ and $(a, f(a))$. If the derivative exists, then the limiting value becomes *slope of the tangent line* to the graph at $x = a$. If the function $f$ is differentiable at $x = a$, then the *linear approximation* to $f$ at $a$ is the function

$$L(x) = f(a) + f'(a)(x - a).$$

For example, if $f(x) = x^2$, then linear approximation to this function at $a = 2$ is $L(x) = f(2) + f'(2)(x - 2) = 4 + 4(x - 2) = 4x - 4$. The following second characterization of differentiability, in terms of linear approximation, is useful in understanding differentiability in higher dimensions.

**Proposition 11.** *Let $f : \mathbb{R} \to \mathbb{R}$. Then $f$ is differentiable if and only if there is a function $L(x) := mx$ such that*

$$\lim_{h\to 0} \frac{|f(a + h) - f(a) - L(h)|}{|h|} = 0.$$

Applying the sequential criterion for limit, we can obtain the following *sequential characterization* of derivatives. This characterization is often useful when trying to show that a given function is not differentiable at a particular point.

**Proposition 12.** *Let $S$ be an open set containing the point $a$, and suppose $f : S \to \mathbb{R}$. Then $f$ is differentiable at $a$ if and only if for every sequence $\{x_n\} \in S$ that converges to $a$ with $x_n \neq a$ for all $n$, the sequence*

$$\left\{ \frac{f(x_n) - f(a)}{x_n - a} \right\}$$

*converges. Furthermore, if $f$ is differentiable at $a$, then the above sequence of quotients converges to $f'(a)$.*

The proofs of the above two propositions are left to the reader.

## Rules of Differentiation

The familiar rules of differentiation can be easily deduced from the definition that

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}.$$

- The derivative of the sum of two differentiable functions is the sum of their derivatives.
- The derivative of the product $f(x)g(x)$ is equal to $f'(x)g(x) + f(x)g'(x)$ (*Product Rule*).
- The quotient $\dfrac{f(x)}{g(x)}$ has derivative $\dfrac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}$ whenever $g(x) \neq 0$ (*Quotient Rule*).
- The *Chain Rule* for differentiation of the composition of $h = f \circ g$ or $h(x) = f(g(x))$ is more subtle. It turns out that $h'(x) = f'(g(x))\, g'(x)$.

Proofs of the above rules can be found in any calculus book.

## Properties of Differentiable Functions

The first important property is that differentiability implies continuity.

**Theorem 44.** *If $f$ is differentiable at $a$, then $f$ is continuous at $a$.*

*Proof.* To show $f$ is continuous at $a$, we need to show $\lim_{x \to a} f(x) = f(a)$, or equivalently $\lim_{x \to a} (f(x) - f(a)) = 0$. Now consider

$$\begin{aligned} \lim_{x \to a} (f(x) - f(a)) &= \lim_{x \to a} \frac{f(x) - f(a)}{x - a} (x - a) \\ &= \lim_{x \to a} \frac{f(x) - f(a)}{x - a} \lim_{x \to a} (x - a) \\ &= f'(a)\, 0 = 0 \end{aligned} \tag{1.3}$$

as desired. Note that going from the first equality to the second is possible because both limits exist. □

**Remark 24.** Continuity does not imply differentiability. For example, $f(x) = |x|$ is continuous at $x = 0$ but not differentiable there. Since $x \to 0$ implies $|x| \to 0$, $f$ is continuous at 0. On the other hand, notice when $h > 0$, $|h| = h$, but when $h < 0$, $|h| = -h$; thus when we form the difference quotient, we get two different limits as $h \to 0^+$ or $h \to 0^-$:

$$\lim_{h\to 0^+} \frac{f(h) - f(0)}{h} = 1 \quad \text{and} \quad \lim_{h\to 0^-} \frac{f(h) - f(0)}{h} = -1.$$

Since the limit exists if and only if its one-sided limits exist and are equal, the above limit does not exist when $a = 0$, and therefore $f(x) = |x|$ is not differentiable at 0.

**Remark 25.** The relationship between continuity and differentiability has led to many results. Of course as an absolute value function demonstrates a function can be continuous but not differentiable at some point. It is easy to construct a function which fails to be differentiable at a *finite set*. If we go one step further and ask the question is it possible to construct a function that is continuous for all of $\mathbb{R}$ but fails to be differentiable at every rational points? Karl Weierstrass presented an example of a continuous function that is not differentiable at *any* point. For more about Weierstrass function, see Theorem 98 in Chapter 2.

**Example 47.** If a function $f$ is differentiable at a certain subset $I \subset \mathbb{R}$, its derivative $f'(x)$ may or may not be continuous there. For example, the function

$$f(x) = \begin{cases} x^2 \sin(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

is differentiable on $\mathbb{R}$, but $f'(x)$ is not continuous on any interval which contains the origin. For differentiability at $x = 0$,

$$f'(0) = \lim_{x\to 0} \frac{f(x) - f(0)}{x - 0} = \lim_{x\to 0} \frac{x^2 \sin(1/x)}{x - 0} = \lim_{x\to 0} x \sin(1/x),$$

which we have already showed by the squeeze principle is zero. Moreover,

$$f'(x) = 2x \sin(1/x) - \cos(1/x),$$

and $\lim_{x\to 0} f'(x)$ does not exist. A function with a continuous derivative is called a $\mathcal{C}^1$-function. Thus the function given in this example is not a $\mathcal{C}^1$-function.

Note that if a function has continuous derivatives to a certain order, say $k$, we say that function is a $\mathcal{C}^k$ function and infinitely differentiable functions are referred as $\mathcal{C}^\infty$ or *smooth* functions. For example, $f(x) = e^x$ is a smooth function.

If a function $f$ is defined on an open set $S$ containing the point $a$, we say that it has a *local maximum* at $a$ if $f(x) \leq f(a)$ for all $x$ in the neighborhood

$$V_\delta(a) = \{x \in \mathbb{R} : \ |x - a| < \delta\} \subset S.$$

A *local minimum* is defined similarly.

**Theorem 45.** *Let $f$ be a function defined on some open set $S$, and suppose it is differentiable at $a \in S$. If $f$ has a local maximum or a local minimum at $a \in S$, then $f'(a) = 0$.*

*Proof.* Suppose $f$ has a local minimum at $a$. By hypothesis, $f(a) \leq f(x)$ for all $x$ in some neighborhood $V_\delta(a)$. This implies that the difference quotient

$$g(x) = \frac{f(x) - f(a)}{x - a} \quad \text{for} \quad x \neq a$$

defined in $V_\delta(a)$ except at $a$ satisfies the inequalities:

$$g(x) \geq 0 \quad \text{for} \quad x > a$$

$$g(x) \leq 0 \quad \text{for} \quad x < a.$$

Since $f$ is differentiable at $a$, it follows that

$$f'(a) = \lim_{x \to a^+} g(x) \geq 0 \quad \text{and} \quad f'(a) = \lim_{x \to a^-} g(x) \leq 0.$$

Hence $f'(x) = 0$ as claimed. □

## The Mean Value Theorem

The Mean Value Theorem is an important and fundamental property of differentiable functions. It is useful in the same way the Intermediate Value Theorem or the Extreme Value Theorem are to continuous functions. The basic ingredient in the Mean Value Theorem is the following:

**Theorem 46** (Rolle's Theorem)**.** *If a function is continuous in an interval $[a, b]$ and differentiable in the open interval $(a, b)$ and if $f(a) = f(b)$, then $f'(c) = 0$ for some $c \in (a, b)$ (Figure 1.39).*

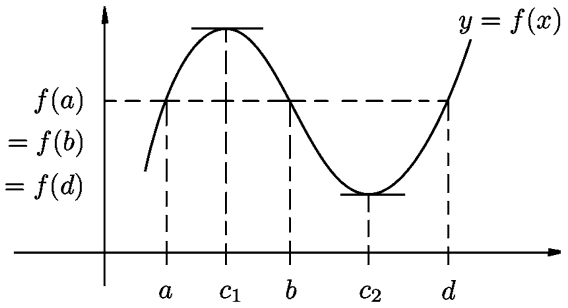



Figure 1.39: Rolle's theorem

*Proof.* Suppose $f(x) = k$ for some constant $k$ on $[a, b]$; then $f'(x) = 0$ in $(a, b)$, and thus there is nothing to prove. Assume $f$ is not constant; then $f$ attains a maximum or a minimum at some point $c \in (a, b)$, and it follows from the previous theorem that $f'(c) = 0$. □

**Theorem 47** (Mean Value Theorem)**.** *If a function $f$ is continuous in an interval $[a, b]$ and differentiable in the open interval $(a, b)$, then (Figure 1.40)*

$$f'(c) = \frac{f(b) - f(a)}{b - a} \quad \text{for some} \quad c \in (a, b).$$

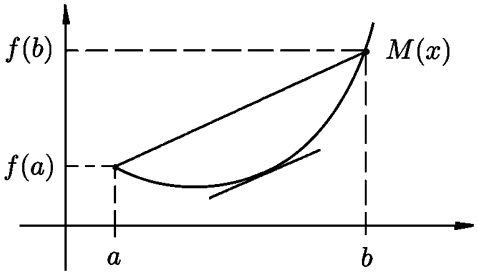



Figure 1.40: The Mean Value Theorem

*Proof.* Define the linear function

$$M(x) = \frac{f(b) - f(a)}{b - a}(x - a) + f(a).$$

Notice that $M(a) = f(a)$ and $M(b) = f(b)$. Then apply Rolle's theorem to the function $g(x) = f(x) - M(x)$. □

The Mean Value Theorem can be applied to show that a function with zero derivative throughout an interval must be a constant function. A more general statement is the following:

**Corollary 10.** *If the derivative of two functions coincides in some interval, the two functions must differ by a constant.*

**Corollary 11.** *A function with positive derivative in an interval is strictly increasing. Similarly, a function with negative derivative is strictly decreasing.*

**Remark 26.** In the following we point out a few things about the Mean Value Theorem. For more about differentiation on $\mathbb{R}$, we refer the reader to [52].

- *The Generalized Mean Value Theorem*, sometimes referred to as *Cauchy's Mean Value Theorem*, states that if $f$ and $g$ are both continuous on $[a, b]$ and differentiable on $(a, b)$, then there is a point $c \in (a, b)$ such that

  $$f'(c)[g(b) - g(a)] = g'(c)[f(b) - f(a)].$$

  This theorem is useful when comparing derivatives of two functions, and it can be proved by setting $h(x) = f(x)[g(b) - g(a)] = g(x)[f(b) - f(a)]$ and applying Rolle's theorem to $h(x)$.

- One can extract information about the function $f$ from $f'$ using the Mean Value Theorem. For example, suppose you are asked to prove the inequality $\ln(1+x) \leq x$ for any $x \geq 0$. First assume $x > 0$. The Mean Value Theorem applied to $\ln(1+x)$ on the interval $[0, x]$ implies the existence of $c \in (0, x)$ such that

$$\frac{\ln(1+x) - \ln(1+0)}{x-0} = \frac{1}{1+c},$$

or $\ln(1+x) = \dfrac{x}{1+c}$. Since $0 < \dfrac{1}{1+c} < 1$, we get $\ln(1+x) < x$. This implies $\ln(1+x) \leq x$ for any $x \geq 0$.

## Inverse Function Theorem

There are various ways of forming new functions from old functions by using addition, multiplication, division, composition, etc. Now we will see another way to construct functions which might double the list of functions we might have. Recall that a function is *one-to-one* or *injective* if $f(a) \neq f(b)$ whenever $a \neq b$. The identity function $I$ is one-to-one, but the function $f(x) = x^2$ is not so, since $f(-1) = f(1)$. However if we restrict the domain of $f(x) = x^2$ to the set of $x \geq 0$, then $f$ becomes one-to-one. Recall also that a function is a collection of pairs of numbers subject to some constraints.

**Definition 35.** For any function $f$, the inverse of $f$, denoted by $f^{-1}$, is the set of all pairs $(a, b)$ for which the pair $(b, a)$ is in $f$.

**Proposition 13.** *$f^{-1}$ is a function if and only if $f$ is one-to-one.*

*Proof.* Suppose $f^{-1}$ is a function. If $f(b) = f(c)$, then $f$ contains the pairs $(b, f(b))$ and $(c, f(c))$ and $(c, f(c)) = (c, f(b))$, so $(f(b), b)$ and $(f(b), c)$ are in $f^{-1}$. Since $f^{-1}$ is a function, this implies that $b = c$. Thus $f$ is one-to-one. Conversely suppose $f$ is one-to-one. Let $(a, b)$ and $(a, c)$ be two pairs in $f^{-1}$. Then $(b, a)$ and $(c, a)$ are in $f$, so $a = f(b)$ and $a = f(c)$, but $f$ is one-to-one, so $f(b) = f(c)$ implies $b = c$. Thus $f^{-1}$ is a function. □

Since the points $(a, b)$ and $(b, a)$ are reflections of each other through the graph of $I(x) = x$ (this is called the *diagonal*), to obtain the graph of $f^{-1}$, one can reflect the graph of $f$ through the line $I(x) = x$ as seen in Figure 1.41.

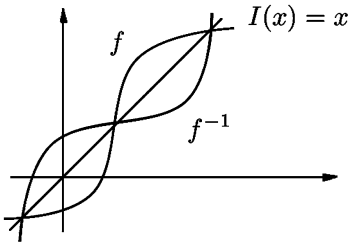



Figure 1.41: The inverse of a function $f$

Note that $f^{-1}(b)$ is the unique number $a$ such that $f(a) = b$; this gives us another way to define the inverse. Observe that

$$f(f^{-1}(x)) = x \quad \text{for all } x \text{ in the domain of } f^{-1} \text{ and}$$

$$f^{-1}(f(x)) = x \quad \text{for all } x \text{ in the domain of } f.$$

These two equations can be expressed as

$$f \circ f^{-1} = I \quad \text{and} \quad f^{-1} \circ f = I.$$

Since many standard functions can be defined as the inverse of other functions, it becomes clear to us that we must identify functions which are one-to-one, and it will be useful to know how the properties of $f$ and $f^{-1}$ are related. It is not difficult to see that if $f$ is increasing, then $f^{-1}$ is also increasing, and if $f$ is decreasing, $f^{-1}$ is decreasing (try to prove it). It is not true that every one-to-one function is either increasing or decreasing as seen in the following example (Figure 1.42):

$$f(x) = \begin{cases} x & \text{if } x \neq 1, 2 \\ 1 & \text{if } x = 2. \\ 2 & \text{if } x = 1. \end{cases}$$

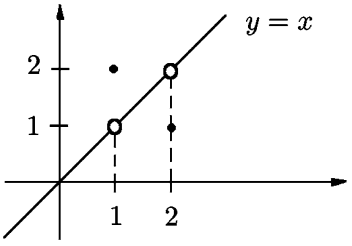



Figure 1.42: A function that is neither increasing nor decreasing

However, every continuous one-to-one function defined on an interval is either increasing or decreasing.

**Theorem 48.** *If $f$ is continuous and one-to-one on a closed and bounded interval $[a, b]$, then $f$ is either increasing or decreasing on that interval.*

For the proof we refer the reader to [52], p. 125. Notice that when $f$ is a continuous and increasing function on a closed interval $[a, b]$, then by the Intermediate Value Theorem $f$ takes every value between $f(a)$ and $f(b)$. Therefore, the domain of $f^{-1}$ is the closed interval $[f(a), f(b)]$. Similarly if $f$ is continuous and decreasing on $[a, b]$, then the domain of $f^{-1}$ is $[f(b), f(a)]$. When using this theorem, you can assume $f$ is increasing and use the standard trick of considering $-f$ to move to decreasing functions.

**Proposition 14.** *If $f$ is continuous and one-to-one on an interval, then $f^{-1}$ is also continuous.*

*Proof.* We know by the previous theorem that $f$ is either increasing or decreasing. Without loss of generality assume that $f$ is increasing. To show $f^{-1}$ is continuous, we must show for every $b$ in the domain of $f^{-1}$ we have $\lim_{x\to b} f^{-1}(x) = f^{-1}(b)$. Note that such a number $b$ has the form $f(a)$; thus using the $\epsilon - \delta$ definition of the limit, for any $\epsilon > 0$, we want to find $\delta$ such that

$$\text{if} \quad |x - f(a)| < \delta \quad \text{then} \quad |f^{-1}(x) - a| < \epsilon.$$

Now since $a \in (a-\epsilon, a+\epsilon)$, it follows that $f(a) \in (f(a-\epsilon), f(a+\epsilon))$. Thus the choice of

$$\delta = \min\{f(a+\epsilon) - f(a), f(a) - f(a-\epsilon)\}$$

works (see Figure 1.43).

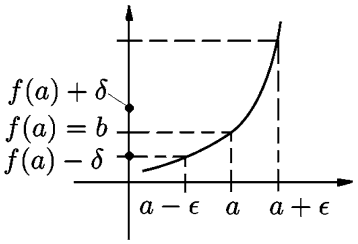



Figure 1.43: Continuity of $f^{-1}$

Our choice ensures that $f(a-\epsilon) \le f(a)-\delta$ and $f(a)+\delta \le f(a+\epsilon)$; therefore $|x-f(a)| < \delta$ implies $f(a-\epsilon) < x < f(a+\epsilon)$, but $f^{-1}$ is an increasing function because $f$ is increasing. Thus we have

$$f^{-1}\left(f(a-\epsilon)\right) < f^{-1}(x) < f^{-1}\left(f(a+\epsilon)\right)$$

or equivalently $|f^{-1}(x) - a| < \epsilon$, which is what we wanted to show. □

Now that we established the continuity of $f^{-1}$, the next step is to investigate the differentiability of $f^{-1}$. Again the following picture gives us the hope that tangent line $L$ at the point $(a, f(a))$ to a one-to-one function $f$ should be related nicely to the tangent line $L^*$ at the point $(f(a), a)$ to the function $f^{-1}$ (Figure 1.44).

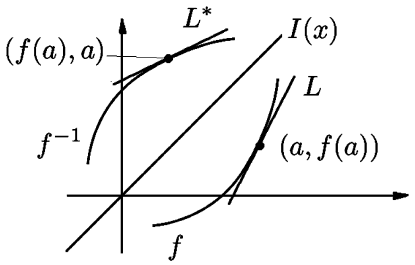



Figure 1.44: Tangent lines to $f$ and $f^{-1}$

If we rush and apply the Chain Rule to $f\left(f^{-1}(x)\right) = x$ by differentiating both sides, we get

$$f'\left(f^{-1}(x)\right)(f^{-1})'(x) = 1,$$

so

$$(f^{-1})'(x) = \frac{1}{f'\left(f^{-1}(x)\right)}.$$

But this is not a proof that $f^{-1}$ is differentiable; we cannot even apply the Chain Rule unless $f^{-1}$ is already known to be differentiable. But suppose that $f^{-1}$ is differentiable, then from above we know what the derivative $(f^{-1})'(x)$ is going to look like. This observation also teaches us that even if $f$ is a continuous one-to-one function defined on an interval and if $f'\left(f^{-1}(a)\right) = 0$, then $f^{-1}$ is not differentiable at $a$. Let $y = f(x)$ be a $\mathcal{C}^1$ -function and $f'(a) \neq 0$; then the following theorem asserts that locally near $a$ we can solve for $x$ to give an inverse function $x = f^{-1}(y)$. We also learn that since $f'(y) = \dfrac{1}{f'(x)}$, it is possible that $y = f(x)$ can be inverted because $f'(a) \neq 0$ means that the slope of $y = f(x)$ is nonzero, so that the graph is rising or falling near $a$. Thus if we reflect the graph across the line $y = x$, it is still a graph near $(a, b)$, where $b = f(a)$. In the following figure we can invert $y = f(x)$ in the shaded box, so that in this range $x = f^{-1}(y)$ is defined. In summary, if $f'(a) \neq 0$, then $y = f(x)$ is locally invertible (Figure 1.45).

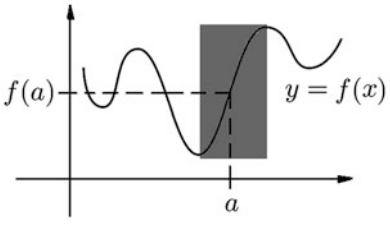



Figure 1.45: Local invertibility

Now we are ready to state the Inverse Function Theorem.

**Theorem 49** (Inverse Function Theorem)**.** *Let $f$ be a continuous one-to-one function defined on an interval, and suppose that $f$ is differentiable at $f^{-1}(b)$, with derivative $f'\left(f^{-1}(b)\right) \neq 0$. Then $f^{-1}$ is differentiable at $b$, and*

$$(f^{-1})'(b) = \frac{1}{f'\left(f^{-1}(b)\right)}.$$

*Proof.* Since we have a continuous one-to-one function $f$, by the above theorems we deduce that $f$ is monotone, and $f^{-1}$ exists and is continuous. Let $b = f(a)$, then $f$ differentiable at $f^{-1}(b)$ ensures that the limit

$$f'(a) = \lim_{k\to 0} \frac{f(a+k) - f(a)}{k}$$

exists. Moreover, by hypothesis, $f'(a) \neq 0$. Now consider the limit

$$\lim_{h\to 0} \frac{f^{-1}(b+h) - f^{-1}(b)}{h}.$$

Since $b + h$ is a number in the domain of $f^{-1}$, it can be written as $f(a + k)$ for some $k = k(h)$, and rewriting the limit

$$\lim_{h\to 0} \frac{f^{-1}(b+h) - f^{-1}(b)}{h} = \lim_{h\to 0} \frac{f^{-1}(f(a+k)) - a}{h} = \lim_{k\to 0} \frac{k}{f(a+k) - b}$$

provided that $k \to 0$ as $h \to 0$ too. This is not difficult to show and follows from the continuity of $f^{-1}$ at $b$, i.e., $\lim_{h\to 0} f^{-1}(b + h) = f^{-1}(b)$. Now we set

$$b + h = f(a + k), \quad \text{thus} \quad k = f^{-1}(b + h) - f^{-1}(b).$$

Since

$$\lim_{k\to 0} \frac{f(a+k) - f(a)}{k} = f'(a) = f'(f^{-1}(b)) \neq 0,$$

we have that

$$(f^{-1})'(b) = \frac{1}{f'\left(f^{-1}(b)\right)}.$$ □

**Example 48.**

a) Let $f(x) = x^3$. Since $f'(x) = 3x^2$ and thus $f'(0) = 0$, the inverse function $f^{-1}(x) = \sqrt[3]{x}$ is not differentiable at 0. Note that this example illustrates the necessity of the condition $f'\left(f^{-1}(b)\right) \neq 0$ in the above theorem.

b) Let $f : [0, \infty) \to \mathbb{R}$ be defined as $f(x) = x^2 + 2$. Then the function $f^{-1}(x) : [2, \infty) \to \mathbb{R}$ defined by $f^{-1}(x) = \sqrt{x - 2}$ is the inverse function, since

$$f(f^{-1}(x)) = \left(\sqrt{x-2}\right)^2 + 2 = x - 2 + 2 = x \quad \text{for all } x \in [2, \infty)$$

and

$$f^{-1}(f(x)) = \sqrt{(x^2+2)-2} = \sqrt{x^2} = x \quad \text{for all} \quad x \in [0,\infty).$$

Note that $f(2) = 6$ and $f'(x) = 2x$, so $f'(2) = 4$. By the Inverse Function Theorem,

$$(f^{-1})'(6) = \frac{1}{f'(2)} = \frac{1}{4}.$$

We can check this by finding $(f^{-1})'(x) = \frac{1}{2}(x-2)^{-1/2}$, so $(f^{-1})'(6) = 1/4$.

c) For $n$ odd let $f(x) = x^n$ for all $x$, and for $n$ even, let $f(x) = x^n$ for $x \geq 0$. Then $f$ is a continuous and one-to-one function whose inverse function is $f^{-1}(x) = \sqrt[n]{x}$. By the Inverse Function Theorem, we have for $x \neq 0$

$$(f^{-1})'(x) = \frac{1}{f'\left(f^{-1}(x)\right)} = \frac{1}{n(x^{1/n})^{n-1}} = \frac{1}{n}x^{(1/n)-1}.$$

What we have shown is if one considers a function $f(x) = x^a$, where $a$ is an integer or the reciprocal of a natural number, then the derivative is given by $f'(x) = ax^{a-1}$. Now you can check that this formula is also true if $a$ is a rational number of the form $a = \frac{m}{n}$.

**Remark 27.** There is a wonderful calculus book by Spivak [46], pp. 200–210, which contains a detailed discussion of inverse functions. See also Section 1.6 for continuous functions on compact sets and Section 2.1 for metric spaces. It turns out that if $f$ is a continuous and one-to-one mapping of a compact metric space $M$ onto a metric space $N$, then the inverse mapping $f^{-1}$ can be defined on $N$ by $f^{-1}(f(x)) = x$ for all $x \in M$ and $f^{-1}$ is a continuous mapping of $N$ onto $M$ (see, e.g., [42], p. 221). For the Inverse Function Theorem for a $\mathcal{C}^1$-function $f : A \subset \mathbb{R}^n \to \mathbb{R}^n$ (a function of several variables), we refer the reader to [26], p. 392.

## Exercises

1. Discuss the differentiability of $f(x) = |x^2 - 9|$ at $x = 2$.

2. Determine whether or not $f$ is differentiable at 0.

   a) $f(x) = \sqrt[3]{x}$.

   b) $f(x) = \sqrt{|x|}$.

   c) $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x \sin\left(\frac{1}{x}\right)$ if $x \neq 0$ and $f(0) = 0$.

   d) $f(x) = x|x|$.

3. Let $f$ be a function on $[a, b]$ that is differentiable at $c$. Let $L(x)$ be the tangent line to $f$ at $c$. Prove that $L$ is the unique linear function with the property that

$$\lim_{x \to c} \frac{f(x) - L(x)}{x - c} = 0.$$

4. Find the derivative (if it exists) of

$$f(x) = \begin{cases} x^2 e^{-x^2} & \text{if } |x| \leq 1 \\ \dfrac{1}{e} & \text{if } |x| > 1. \end{cases}$$

5. We say a function $f : (a, b) \to \mathbb{R}$ is uniformly differentiable if $f$ is differentiable on $(a, b)$, and for each $\epsilon > 0$ there exists a $\delta > 0$ such that

$$0 < |x - y| < \delta \quad \text{and} \quad x, y \in (a, b) \quad \Rightarrow \quad \left| \frac{f(x) - f(y)}{x - y} - f'(x) \right| < \epsilon.$$

Prove that if $f$ is uniformly differentiable, then $f'$ is continuous. Then give an example of a function that is differentiable but not uniformly differentiable.

6. Let $f : \mathbb{R} \to \mathbb{R}$. Assume that for any $x, t \in \mathbb{R}$, we have

$$|f(x) - f(t)| \leq |x - t|^{1+\alpha},$$

where $\alpha > 0$. Show that $f(x)$ is constant.

7. Does the Mean Value Theorem apply to $f(x) = \sqrt{x}$ on $[0, 1]$? Does it apply to $g(x) = \sqrt{|x|}$ on $[-1, 1]$?

8. Consider the function

$$f(x) = \begin{cases} e^{-1/x^2} & \text{if } x \neq 0 \\ \\ 0 & \text{if } x = 0. \end{cases}$$

Show that the $n$th derivative $f^{(n)}$ at 0 is $f^{(n)}(0) = 0$, for $n = 1, 2, \ldots$.

9. Suppose $f$ and $g$ are real functions of a real variable whose $n$th derivatives $f^{(n)}$ exist at a point $a$, where $n \in \mathbb{N}$. Prove the Leibniz generalization of the product formula:

$$(fg)^{(n)}(a) = \sum_{k=0}^{n} \binom{n}{k} f^{(k)}(a) g^{(n-k)}(a).$$

10. For $x \in \mathbb{R}$, let $f$ be defined as $f(x) = x^2 e^{x^2}$. Show that $f^{-1}$ exists and is differentiable on $(0, \infty)$. Compute its derivative $(f^{-1})'(e)$.

11. Suppose $f$ and $g$ are one-to-one continuous functions on $\mathbb{R}$. Compute the derivatives:

a) $(f^{-1})'(2)$

b) $(g^{-1})'(2)$

c) $(f^{-1} \cdot g^{-1})'(2)$

provided that $f(0) = 2,\ g(1) = 2,\ f'(0) = \pi$, and $g'(1) = e$.

12. If $f$ is a continuous one-to-one function defined on an interval and

$$f'(f^{-1}(a)) = 0,$$

then show that $f^{-1}$ is not differentiable at $a$.

13. Let $f_n : \mathbb{R} \to \mathbb{R}$ be differentiable for each $n = 1, 2, \ldots$ with $|f_n'(x)| \leq 1$ for all $n$ and $x$. Assume $\lim_{n\to\infty} f_n(x) = h(x)$ for all $x$. Prove that $h : \mathbb{R} \to \mathbb{R}$ is continuous.

## 1.8 The Riemann Integral

Later we will define the integral in a formal way, but loosely speaking, the integral of a positive function $f(x) \geq 0$ for all $x \in [a, b]$ is just the *area* under its graph. Our aim is to make this notion precise so that we can extend it to a wider class of functions. Even though the germ of the idea of integration goes back to Archimedes in the third century B.C. and we see ideas of integration developed in the seventeen century by I. Newton and G. Leibniz, it was B. Riemann (1826–1866) who gave his definition of an integral using *Riemann sums*. The approach in terms of upper and lower sums is due to G. Darboux (1842–1917). Subsequently, other more general approaches to integration were given by T. J. Stieltjes (1856–1894) and H. Lebesgue (1875–1941). In the following we consider bounded functions $f$ defined on $[a, b]$, meaning that there is a constant $M > 0$ such that $|f(x)| \leq M$ for all $x \in [a, b]$.

**Definition 36.** Let $f$ be a function defined on a bounded interval $[a, b]$. A *partition* $P$ of $[a, b]$ is a finite set of points in $[a, b]$ such that

$$a = x_0 < x_1 < x_2 < \cdots < x_n = b.$$

For $k = 1, 2, \ldots, n$, let

$$M_k = \sup\{f(x) :\ x_{k-1} \leq x \leq x_k\} \quad \text{and} \quad m_k = \inf\{f(x) :\ x_{k-1} \leq x \leq x_k\}.$$

Then by *upper sum*, denoted by $U(f,P)$, and *lower sum*, denoted by $L(f,P)$, we mean

$$U(f,P)=\sum_{k=1}^{n} M_k(x_k-x_{k-1}) \quad \text{and} \quad L(f,P)=\sum_{k=1}^{n} m_k(x_k-x_{k-1}).$$

Clearly $L(f,P)\le U(f,P)$ for every partition $P$. It helps to think of upper sums as an overestimate and lower sums as an underestimate for the value of the integral. If $P$ and $Q$ are two partitions of $[a,b]$ with $P\subset Q$, then $Q$ is called a refinement of $P$. It can be seen that $L(f,P)\le L(f,Q)$ and $U(f,P)\ge U(f,Q)$ (Figure 1.46).

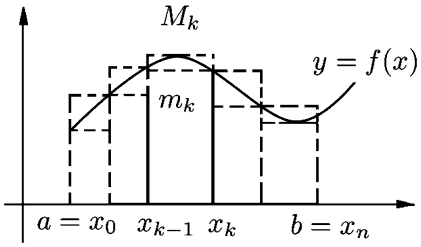



Figure 1.46: Upper and lower sums

Let $\mathcal{P}$ be the collection of all possible partitions of the interval $[a,b]$, and let

$$U(f)=\inf\{U(f,P):\ P\in\mathcal{P}\} \quad \text{and} \quad L(f)=\sup\{L(f,P):\ P\in\mathcal{P}\}.$$

Let $f$ be a bounded function on $[a,b]$ and $P$ be any partition of $[a,b]$. First observe that $U(f)\ge L(f,P)$; then one can prove that for any bounded function $f$ on $[a,b]$ we always have $L(f)\le U(f)$.

**Definition 37.** A bounded function $f$ on $[a,b]$ is *Riemann integrable* over $[a,b]$ if $U(f)=L(f)$ and its integral is defined to be the common value, written as

$$\int_a^b f(x)\,dx=U(f)=L(f).$$

We can also say that $f$ is Riemann integrable if and only if for each $\epsilon>0$, there exists a partition $P$ such that $U(f,P)-L(f,P)<\epsilon$, that is, integrability is equivalent to the existence of partitions whose upper and lower sums are arbitrarily close together (see Proposition 15 below).

**Example 49.** Consider Dirichlet's function

$$f(x)=\begin{cases}1 & \text{if } x\in\mathbb{Q}\\ 0 & \text{if } x\notin\mathbb{Q}.\end{cases}$$

We have seen that this function is discontinuous at every point of $[0,1]$. Now we turn our attention to integrability. If $P$ is some partition of $[a,b]$, then the

density of rationals in $\mathbb{R}$ implies that every subinterval of $P$ contains points for which $f(x) = 1$, and thus $U(f, P) = b - a$. Similarly $L(f, P) = 0$ because irrational numbers are also dense in $\mathbb{R}$. Because this is the case for every possible partition, $U(f) = b - a$ and $L(f) = 0$. The two are not equal, and therefore Dirichlet's function is not Riemann integrable. This example also illustrates that not every bounded function is Riemann integrable.

**Proposition 15** (Riemann's Criterion)**.** *A bounded real-valued function $f$ on $[a, b]$ is Riemann integrable if and only if for each $\epsilon > 0$ there is a partition $P$ such that*

$$U(f, P) - L(f, P) < \epsilon.$$

*Proof.* Suppose $f$ is Riemann integrable on $[a, b]$. Given $\epsilon > 0$, it follows from the definition of upper and lower sums that there are partitions $P_1$ and $P_2$ such that

$$\left(\int_a^b f\,dx\right) - \frac{\epsilon}{2} < L(f, P_1) \quad \text{and} \quad U(f, P_2) < \left(\int_a^b f\,dx\right) + \frac{\epsilon}{2}.$$

Let $P$ be a refinement of $P_1$ and of $P_2$. Then

$$\left(\int_a^b f\,dx\right) - \frac{\epsilon}{2} < L(f, P_1) \le L(f, P) \le U(f, P) \le U(f, P_2) < \left(\int_a^b f\,dx\right) + \frac{\epsilon}{2},$$

which implies $U(f, P) - L(f, P) < \epsilon$. For the proof of the converse statement, let $\epsilon > 0$. If a partition $P*$ of $[a, b]$ exists such that $U(f, P*) - L(f, P*) < \epsilon$, then $U(f) - L(f) \le U(f, P*) - L(f, P*) < \epsilon$. Since $\epsilon$ is arbitrary, we have $U(f) = L(f)$. □

**Example 50.** Consider the following function $f$ defined on the interval $[0, 2]$ (Figure 1.47):

$$f(x) = \begin{cases} 1 & \text{if } x = 1 \\ 0 & \text{if } x \neq 1. \end{cases}$$

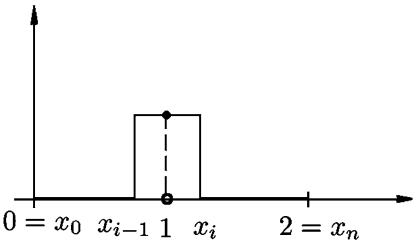



Figure 1.47: Riemann integrable discontinuous function

If $P = \{0 = x_0 < x_1 < x_2 < \cdots < x_n = 2\}$ is a partition of $[0, 2]$ with $x_{i-1} < 1 < x_i$, then

$$m_k = M_k = 0 \quad \text{if} \quad k \neq i \quad \text{but} \quad m_i = 0 \quad \text{and} \quad M_i = 1.$$

Writing upper and lower sums,

$$U(f,P) = \sum_{k=1}^{i-1} M_k(x_k - x_{k-1}) + M_i(x_i - x_{i-1}) + \sum_{k=i+1}^{n} M_k(x_k - x_{k-1})$$

$$L(f,P) = \sum_{k=1}^{i-1} m_k(x_k - x_{k-1}) + m_i(x_i - x_{i-1}) + \sum_{k=i+1}^{n} m_k(x_k - x_{k-1}),$$

which implies

$$U(f,P) - L(f,P) = x_i - x_{i-1}.$$

We can take $|x_i - x_{i-1}| < \epsilon$ for a suitable $\epsilon$ so that

$$U(f,P) - L(f,P) < \epsilon.$$

By Riemann's criterion $f$ is integrable. Clearly this function is discontinuous at $x = 1$. To show $f$ is Riemann integrable, we embedded the point $x = 1$ into a very small subinterval whose length is less than $\epsilon$.

Going back to the definitions of upper and lower sums,

$$U(f,P) - L(f,P) = \sum_{k=1}^{n} (M_k - m_k)(x_k - x_{k-1}),$$

where $M_k$ and $m_k$ are the supremum and infimum of the function over the interval $[x_{k-1}, x_k]$. Said in another way, $M_k - m_k$ is the variation when the domain is restricted to $[x_{k-1}, x_k]$. Thus we expect the following relationship between continuity and integrability.

**Proposition 16.** *If $f$ is continuous on $[a,b]$, then $f$ is Riemann integrable on $[a,b]$.*

*Proof.* To see this we apply the theorem that every continuous function on a closed and bounded (thus compact set) set is uniformly continuous there, meaning, given $\epsilon > 0$, there exists $\delta > 0$ such that

$$|f(x) - f(y)| < \frac{\epsilon}{b-a} \quad \text{whenever} \quad |x - y| < \delta.$$

Now choose a partition $P$ so that $x_k - x_{k-1}$ is less than $\delta$. Also observe that on each compact subinterval $[x_{k-1}, x_k]$, $f$ being a continuous function means it assumes its infimum and supremum at some point there. By the Extreme Value Theorem, there exists $t_k, w_k \in [x_{k-1}, x_k]$, such that $M_k = f(t_k)$ and $m_k = f(w_k)$. But we also have $|t_k - w_k| < \delta$, therefore

$$M_k - m_k < \frac{\epsilon}{b-a},$$

and so

$$U(f,P) - L(f,P) = \sum_{k=1}^{n} (M_k - m_k)(x_k - x_{k-1}) < \epsilon,$$

which shows that $f$ is integrable on $[a,b]$. □

**Remark 28.** The argument used above can be modified to show that if $f$ is a monotonic bounded function, then it is integrable. Because $f$ is bounded, there exists $M > 0$ such that $|f(x)| \leq M$ on $[a, b]$. Given $\epsilon > 0$, choose a partition $P$ of $n$ intervals of length less than $\dfrac{\epsilon}{2M}$. Then

$$U(f, P) - L(f, P) = \sum_{k=1}^{n} (f(x_k) - f(x_{k-1}))(x_k - x_{k-1}) < \frac{\epsilon}{2M}[f(b) - f(a)] \leq \epsilon.$$

Hence $f$ is integrable on $[a, b]$.

## Algebraic Properties of the Integral

Suppose $f$ and $g$ are integrable functions on $[a, b]$, and suppose that $c$ is a constant. Then the functions $cf$, $f + g$, $f \pm g$, and $|f|$ are integrable on $[a, b]$ and

- $\displaystyle\int_a^b cf(x)\,dx = c\int_a^b f(x)\,dx.$
- $\displaystyle\int_a^b [f(x) \pm g(x)]\,dx = \int_a^b f(x)\,dx \pm \int_a^b g(x)\,dx$ (additivity).
- $\displaystyle\left|\int_a^b f(x)\,dx\right| \leq \int_a^b |f(x)|\,dx.$

Moreover, if $f(x) \leq g(x)$ for all $x \in [a, b]$, then $\displaystyle\int_a^b f(x)\,dx \leq \int_a^b g(x)\,dx$. The above algebraic properties follow fairly readily from the definitions. Note also that additivity of the integral suggests an appropriate extension of integrating complex-valued functions. More precisely if $f$ is a bounded complex-valued function on $[a, b]$, meaning if $f(x) = \mathrm{Re} f(x) + i\,\mathrm{Im} f(x)$, then

$$\int_a^b f(x)\,dx = \int_a^b \mathrm{Re} f(x)\,dx + i\int_a^b \mathrm{Im} f(x)\,dx.$$

There is another kind of additivity of the integral stated below, which can be proven by using partitions that include the point $b$.

**Proposition 17.** *Suppose that $a < b < c$. A bounded function $f$ on the interval $[a, c]$ is Riemann integrable on $[a, c]$ if and only if it is Riemann integrable on each of the subintervals $[a, b]$ and $[b, c]$. In this case*

$$\int_a^c f(x)\,dx = \int_a^b f(x)\,dx + \int_b^c f(x)\,dx.$$

The above identity is quite useful especially if we drop the condition $a < b < c$. Assume $f$ is Riemann integrable on $[a, b]$, and define $\displaystyle\int_c^c f(x)\,dx = 0$ for

$c \in [a, b]$; then

$$\int_b^a f(x)\,dx = -\int_a^b f(x)\,dx.$$

The conclusion of the above proposition is valid for any triple of numbers $a, b$, and $c$ for which all three integrals are defined.

## Intermediate Value Theorem for Integrals

The Intermediate Value Theorem that we have seen for continuous functions can be used to obtain a corresponding theorem for integrals.

**Theorem 50.** *If a function $f$ is continuous on an interval $[a, b]$, then (Figure 1.48)*

$$\frac{1}{b-a}\int_a^b f(x)\,dx = f(c) \quad \textit{for some } c \in [a, b].$$

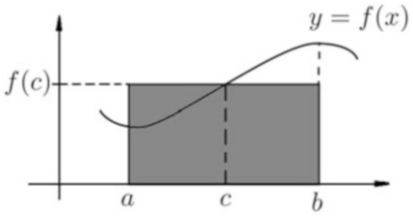



Figure 1.48: Intermediate Value Theorem for integrals

The area of the rectangle with height $f(c)$ and base $(b - a)$ equals the area under $f(x)$ over $[a, b]$.

*Proof.* Since $f$ is continuous on the compact set $[a, b]$, it attains its minimum value $m$ and a maximum value $M$ on the interval $[a, b]$. Since $m \le f(x) \le M$ for all $x \in [a, b]$, it follows that

$$m \le \frac{1}{b-a}\int_a^b f(x)\,dx \le M.$$

An application of Intermediate Value Theorem for continuous functions gives the desired result. □

**Remark 29.** To extend the Riemann integral to unbounded intervals or unbounded functions, we define

$$\int_a^\infty f(x)\,dx = \lim_{b\to\infty}\int_a^b f(x)\,dx \quad \text{and} \quad \int_{-\infty}^b f(x)\,dx = \lim_{a\to-\infty}\int_a^b f(x)\,dx,$$

provided that the limits exist. For example,

$$\int_1^\infty x^{-2}dx = \lim_{b\to\infty}\int_1^b x^{-2}dx = \lim_{b\to\infty}\left.\frac{-1}{x}\right|_1^b = \lim_{b\to\infty}(1-1/b) = 1.$$

Similarly, integrals with a singularity at an end point of integration can be computed. For example, if $f$ is defined in $(a,b]$ and is bounded and integrable on each subinterval $[c,b]$ for $a<c<b$, we define

$$\int_a^b f(x)\,dx = \lim_{c\to a^+}\int_c^b f(x)\,dx,$$

if the limit exists. For example, suppose we have $\displaystyle\int_0^1 \frac{1}{x}\,dx$. Then for each $c\in(0,1)$, we have $\displaystyle\int_c^1 \frac{1}{x}\,dx = \ln 1 - \ln c = -\ln c$. Since $\lim_{c\to 0^+}(-\ln c)=\infty$, the given integral $\displaystyle\int_0^1 \frac{1}{x}\,dx$ diverges. On the other hand, $f(x)=\dfrac{1}{\sqrt{x}}$ is still undefined at $x=0$ but

$$\int_0^1 \frac{1}{\sqrt{x}}\,dx = \lim_{c\to 0^+}\left.2\sqrt{x}\right|_c^1 = 2.$$

## The Fundamental Theorem of Calculus

The derivative and integral are defined independently. The definition of derivative is motivated by the problem of finding slope of the tangent line and is expressed as

$$f'(x) = \lim_{h\to 0}\frac{f(x+h)-f(x)}{h}.$$

It is a functional limit of difference quotients. The definition of integral is motivated by finding areas under non-constant functions and is given in terms of supremums and infimums of finite sums:

$$\int_a^b f(x)\,dx = U(f) = L(f).$$

The Fundamental Theorem of Calculus reveals a remarkable relationship between these two concepts, which says that integration and differentiation are inverse processes. The Fundamental Theorem of Calculus is really two theorems. The first one states under some conditions on the function "the derivative of an integral of a function is the original function." The second result establishes the reverse "the integral of the derivative of a function is again the same function." We start with a theorem about differentiation of the integral.

**Theorem 51** (Fundamental Theorem of Calculus I)**.** *Suppose that $f$ is an integrable function on $[a,b]$, and define*

$$F(x) = \int_a^x f(t)\,dt, \quad a \le x \le b.$$

*Then $F$ is continuous on $[a,b]$. Furthermore, $F$ is differentiable at each point $x_0 \in (a,b)$, where $f$ is continuous, and $F'(x_0) = f(x_0)$.*

*Proof.* By definition $f$ is bounded. Let $M > 0$ be such that $|f(x)| \le M$ for all $x \in [a,b]$. Take $x, y \in [a,b]$, and observe that

$$|F(y) - F(x)| = \left| \int_x^y f(t)\,dt \right| \le M|x-y|.$$

This shows that $F$ is (uniformly) continuous on $[a,b]$. Now let us assume that $f$ is continuous at $x_0 \in [a,b]$. In order to show $F'(x_0) = f(x_0)$, write $F'(x_0)$ as

$$\begin{aligned} \lim_{x\to x_0} \frac{F(x)-F(x_0)}{x-x_0} &= \lim_{x\to x_0} \frac{1}{x-x_0}\left(\int_a^x f(t)\,dt - \int_a^{x_0} f(t)\,dt\right) \\ &= \lim_{x\to x_0} \frac{1}{x-x_0}\left(\int_{x_0}^x f(t)\,dt\right) \end{aligned}$$

and thus

$$\frac{F(x)-F(x_0)}{x-x_0} - f(x_0) = \frac{1}{x-x_0}\int_{x_0}^x [f(t)-f(x_0)]dt, \quad x \in (a,b), x \ne x_0.$$

The assumption of continuity of $f$ at $x_0$ gives us control over the difference $|f(t)-f(x_0)|$, i.e., given $\epsilon > 0$, there is a $\delta > 0$ such that

$$|x-x_0| < \delta \quad \text{implies} \quad |f(x)-f(x_0)| < \epsilon.$$

Therefore,

$$\left| \frac{F(x)-F(x_0)}{x-x_0} - f(x_0) \right| < \epsilon \quad \text{if} \quad 0 < |x-x_0| < \delta,$$

which proves that $F$ is differentiable at $x_0$ and that $F'(x_0) = f(x_0)$. □

Note that the conclusion of the above theorem can be expressed using Leibnitz's notation as

$$\frac{d}{dx}\left(\int_a^x f(t)\,dt\right) = f(x).$$

For example, let $F(x) = \displaystyle\int_0^x \sqrt{t^2+3}\ dt$, for $x \ge 0$, then $F'(x) = \sqrt{x^2+3}$ or

$$\frac{d}{dx}\left(\int_0^x \sqrt{t^2+3}\,dt\right) = \sqrt{x^2+3}.$$

**Example 51.** Let $F(x) = \int_0^{x^2} \sqrt{t^4+3}\, dt$, for $x \geq 0$, and suppose we want to find $F'(x)$ using the Fundamental Theorem of Calculus I. In this problem $F(x) = \int_a^{g(x)} f\, dt$, where $g(x) = x^2$. Thus $F = G \circ g$, where $G(x) = \int_a^x f\, dt$, and we can use the Chain Rule to write

$$F'(x) = G'(g(x))g'(x),$$

where $g'(x) = 2x$ and $G'(g(x)) = \sqrt{(g(x))^4+3}$. Thus

$$F'(x) = \left(\sqrt{(x^2)^4+3}\right)(2x).$$

**Theorem 52** (Fundamental Theorem of Calculus II)**.** *If $f$ is differentiable on $[a,b]$ and $f'$ is integrable on $[a,b]$, then*

$$\int_a^b f' = f(b) - f(a).$$

*Proof.* Let $P$ be a partition of $[a,b]$. By applying the Mean Value Theorem to each subinterval $[x_{k-1}, x_k]$, we obtain points $t_k \in (x_{k-1}, x_k)$ such that

$$f(x_k) - f(x_{k-1}) = f'(t_k)(x_{k-1} - x_k).$$

Thus we have

$$f(b) - f(a) = \sum_{k=1}^{n} [f(x_k) - f(x_{k-1})] = \sum_{k=1}^{n} f'(t_k)(x_{k-1} - x_k),$$

and it follows that

$$L(f', P) \leq f(b) - f(a) \leq U(f', P).$$

Since this holds for each partition $P$, we also have

$$L(f') \leq f(b) - f(a) \leq U(f').$$

But $f'$ is assumed to be integrable on $[a,b]$, thus, $L(f') = U(f') = \int_a^b f'\, dx$ and $\int_a^b f'\, dx = f(b) - f(a)$. □

The Fundamental Theorem of Calculus II provides the standard device of anti-differentiation to compute the integral. For example, if we let $f(x) = x^3$ for $x \in \mathbb{R}$, then $f'(x) = 3x^2$, and

$$\int_1^2 3x^2\, dx = f(2) - f(1) = 8 - 1 = 7.$$

It is also the basis for a well-known formula for *integration by parts* which we will give in the following example.

**Example 52** (Integration by Parts)**.** Suppose that $f$ and $g$ are differentiable on $[a, b]$ and that $f'$ and $g'$ are integrable on $[a, b]$. Then the following integration by parts formula holds:

$$\int_a^b (fg')\,dx = [f(b)g(b) - f(a)g(a)] - \int_a^b (f'g)\,dx.$$

To see this let $h = fg$, then $h' = f'g + fg'$. Now both $f$ and $g$ are differentiable, hence continuous and therefore integrable on $[a, b]$. Thus $h'$ is integrable on $[a, b]$, and from the Fundamental Theorem of Calculus II, we obtain

$$\int_a^b h'\,dx = h(b) - h(a),$$

that is,

$$\int_a^b [f'g + fg']\,dx = \int_a^b (f'g)\,dx + \int_a^b (fg')\,dx = f(b)g(b) - f(a)g(a).$$

## Sets of Measure Zero and Lebesgue's Criterion

When we constructed the Cantor set, we saw that it has zero length, and we stated this as "the Cantor set has measure zero." Now we give a precise definition to this concept.

**Definition 38.** A set $A \subseteq \mathbb{R}$ is said to be *measure zero* if, for every $\epsilon > 0$, there exists a countable collection of open intervals $O_n$ such that

$$A \subseteq \bigcup_{n=1}^{\infty} O_n \quad \text{and} \quad \sum_{n=1}^{\infty} |O_n| \leq \epsilon,$$

where by $|O_n|$ we mean the length of the interval $O_n$.

It is clear that if $A$ is a set of measure zero and $K \subset A$, then $K$ has measure zero as well.

**Example 53.**

a) A single point $A = \{a\}$ is a set of measure zero. For $n = 1$, let $O_1 = (a - \epsilon, a + \epsilon)$, and for $n \geq 2$, set $O_n = \emptyset$; then $\bigcup_{n=1}^{\infty} O_n$ covers $A$ and the intervals have the total length $\epsilon$.

b) A finite set $A = \{a_1, a_2, \ldots, a_k\}$ has measure zero. Let $\epsilon > 0$ be arbitrary, and for each $1 \leq n \leq k$, consider open sets of the form $O_n = \left(a_n - \dfrac{\epsilon}{2k}, a_n + \dfrac{\epsilon}{2k}\right)$. Then

$$A \subseteq \bigcup_{n=1}^{k} O_n \quad \text{and} \quad \sum_{n=1}^{k} |O_n| = \sum_{n=1}^{k} \frac{\epsilon}{k} = \epsilon.$$

c) If $A$ is a countable subset of $\mathbb{R}$, then $A$ is a set of measure zero. Let $A = \{a_1, a_2, \dots\}$ be a countable infinite set. Let $\epsilon > 0$ be arbitrary and for $i \in \mathbb{N}$; set

$$O_i = \left(a_i - \frac{\epsilon}{2^{i+1}}, a_i + \frac{\epsilon}{2^{i+1}}\right).$$

Observe that $a_i \in O_i$ and $|O_i| = \epsilon 2^{-i}$ for $i \in \mathbb{N}$. Therefore,

$$A \subseteq \bigcup_{i=1}^{\infty} O_i \quad \text{and} \quad \sum_{n=1}^{\infty} |O_i| = \epsilon \sum_{n=1}^{\infty} \frac{1}{2^i} = \epsilon.$$

d) $\mathbb{Q}$ has measure zero since it is countable.

e) If $A_1, A_2, \dots$ is a sequence of sets of measure zero, then $A = \bigcup_{n=1}^{\infty} A_n$ is also a set of measure zero (try to prove it).

f) The Cantor set $C$ has measure zero. Recall that the Cantor set $C$ is uncountable, but for each $n$, the Cantor set is contained in a finite union of intervals of total length $\frac{2^n}{3^n}$ and (measure of $C$) $\leq \frac{2^n}{3^n} \to 0$.

Above we proved that continuous functions are Riemann integrable and saw examples of functions with only a finite number of discontinuities that are also integrable. However the following example illustrates the fact that the set of discontinuities of an integrable function can be infinite, even uncountable.

**Example 54.** Let $0 < a < b$ and

$$f(x) = \begin{cases} \dfrac{1}{q} & \text{if } x \in [a,b] \cap \mathbb{Q} \text{ and} x = \dfrac{p}{q} \text{ where } p \text{ and } q \text{ are coprime;} \\ \\ 0 & \text{if } x \in [a,b] \text{ is irrational.} \end{cases}$$

This function is continuous at each irrational number in $(0,1)$ and discontinuous at each rational number in $(0,1)$. We claim that $f(x)$ is Riemann integrable over $[a,b]$ and that $\int_a^b f(x)dx = 0$.

Let $\mathcal{P}$ be a partition of $[a,b]$. If $\mathcal{P} = \{x_0 = a < x_1 < \cdots < x_n = b\}$, then

$$\inf\{f(x) : x_i < x < x_{i+1}\} = 0$$

because $\mathbb{R} \setminus \mathbb{Q}$ is dense in $\mathbb{R}$. Hence we have $L(\mathcal{P}, f) = 0$, which implies

$$L(f) = \sup L(\mathcal{P}, f) = 0,$$

where the supremum is taken over all partitions of $[a,b]$. Let us now show that $U(f) = L(f) = 0$. Fix $\epsilon > 0$. Then the set

$$B_\epsilon = \left\{ \frac{p}{q} \in [a,b] \cap \mathbb{Q};\ \text{where } p \text{ and } q \text{ are coprime and } \frac{1}{q} \geq \frac{\epsilon}{2(b-a)} \right\}$$

is finite. Without loss of generality, assume that $B_\epsilon$ is not empty and has $n \geq 1$ elements. Set

$$B_\epsilon = \{x_1 < x_2 < \cdots < x_n\} .$$

Assume for now $a < x_1$ and $x_n < b$. Choose $m \geq 1$ large enough to have $\frac{1}{m} < \frac{\epsilon}{2n}$, $x_i + \frac{1}{2m} < x_{i+1} - \frac{1}{2m}$, $a < x_1 - \frac{1}{2m}$, and $x_n + \frac{1}{2m} < b$. Consider the partition

$$\mathcal{P}_0 = \left\{ a, x_i - \frac{1}{2m}, x_i + \frac{1}{2m}, b \right\} .$$

We have

$$0 \leq \sup \left\{ f(x) : \ x_i - \frac{1}{2m} \leq x \leq x_i + \frac{1}{2m} \right\} \leq 1$$

because $0 \leq f(x) \leq 1$ for all $x \in [a, b]$. On any other interval $I$ associated with the partition, we have

$$0 \leq \sup \{ f(x) : \ x \in I \} \leq \frac{\epsilon}{2(b-a)}$$

because $I \cap B_\epsilon$ is empty. Hence

$$0 \leq U(\mathcal{P}_0, f) \leq n\frac{1}{m} + \frac{\epsilon}{2(b-a)}(b-a) = \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon .$$

If $x_1 = a$, then we consider only the interval $[a, a + 1/m]$, and if $x_n = b$, then we consider only the interval $[b - 1/m, b]$. The proof is carried similarly to get $U(\mathcal{P}_0, f) \leq \epsilon$. Clearly this will imply $\inf U(\mathcal{P}, f) = 0$, where the infimum is taken over all partitions of $[a, b]$. Hence

$$L(f) = U(f) = 0 ,$$

which implies that $f(x)$ is Riemann integrable and $\displaystyle\int_a^b f(x)dx = 0$.

**Theorem 53** (Lebesgue's Theorem)**.** *Let $f$ be a bounded function defined on $[a, b]$. Then $f$ is Riemann integrable if and only if the set of points where $f$ is not continuous has measure zero.*

The proof of Lebesgue's theorem depends on making several observations on the set of oscillations of $f$ and finding a partition $P$ so that $U(f) - L(f) < \epsilon$. For the details of the proof, consult [52], p. 473. Using Lebesgue's theorem, we can decide the Riemann integrability of certain functions quite easily. For example, if we have

$$f(x) = \begin{cases} \sin\left(\frac{1}{x}\right) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0, \end{cases}$$

then we can claim $f$ is Riemann integrable on $[-1, 1]$. Note that the length of $[-1, 1] \neq 0$ and $f$ has one point of discontinuity at $x = 0$; thus the set

of discontinuities has measure zero. Additionally $f$ is bounded since $|f(x)| \leq 1$; by the above theorem it follows that it is Riemann integrable. The set of discontinuities for the function in Example 54 is $\mathbb{Q}$. Since $\mathbb{Q}$ is countable, it has measure zero, so we again conclude that $f$ is Riemann integrable.

**Remark 30.** For the Riemann integral we divide the area under the graph of $f$ into vertical rectangles to find upper and lower sums. How about dividing this area into horizontal rectangles? This leads to a more complicated mathematics. Suppose $f : [a, b] \to \mathbb{R}$ is a nonnegative bounded function defined on $[a, b]$. Let $R = \{y_0 < y_1 < \cdots < y_n\}$ be a partition of the range of $f$ as shown in Figure 1.49.

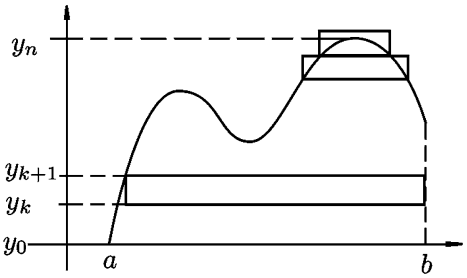



Figure 1.49: Dividing the area under $f$ into horizontal rectangles

Suppose we want to find out

$$\sum_{k=1}^{n}(y_{k+1} - y_k) \text{ (length of the interval } \{x : f(x) \geq y_k\}).$$

However, the set $\{x : f(x) \geq y_k\}$ might be a complicated set, and one asks how can one find its length? To answer this question one has to develop the idea of nonzero measure (or length) of a set. Among many notions of integral, the most prominent one is the Lebesgue integral. To develop these ideas precisely requires a text in itself (see [50]). We return to the idea of Lebesgue integral in Chapter 2, Section 2.7.

## Exercises

1. Let $f(x) = 1 - x^2$. Compute $U(f, P)$, $L(f, P)$ and $\int_0^1 f(x)\,dx$, where

$$P = \left\{0, \frac{2}{5}, \frac{1}{2}, \frac{3}{5}, 1\right\}.$$

2. Suppose $f$ is continuous on $\mathbb{R}$. Explain why the functions defined by $f^3(x)$, $\cos(f(x))$, or $f(\cos x)$ are all integrable over every interval $[a, b]$.

3. Let $f : \mathbb{R} \to \mathbb{R}$ be continuous and set $F(x) = \int_0^{x^2} f(y)\,dy$. Prove that $F'(x) = f(x^2)\,2x$.

4. Suppose $f$ is continuous on a non-degenerate interval $[a, b]$. Show that $\int_a^c f(x)\,dx = 0$ for all $c \in [a, b]$ if and only if $f(x) = 0$ for all $x \in [a, b]$.

5. For $x > 0$, define $L(x) = \int_1^x \frac{1}{t}\,dt$. Prove the following:

a) $L$ is increasing.

b) $L(xy) = L(x) + L(y)$.

c) $L'(x) = \frac{1}{x}$.

d) $L(1) = 0$.

e) What is $L$? (Note that properties c) and d) uniquely determine $L$).

6. We call $U : [a, b] \to \mathbb{R}$ *a step function* if there is a partition $\mathcal{P}$ of $[a, b]$ so that $U$ is constant on each interval of $\mathcal{P}$. Show that any function $f : [a, b] \to \mathbb{R}$ is Riemann integrable if and only if for each $\varepsilon > 0$, there exist two step functions $U$ and $V$ on $[a, b]$ such that $V(x) \le f(x) \le U(x)$ and

$$\int_a^b \Big(U(x) - V(x)\Big)dx < \epsilon.$$

7. Let $f : [a, b] \to \mathbb{R}$ be a Riemann integrable function. Show that $|f(x)|$ is Riemann integrable and

$$\left|\int_a^b f(x)\,dx\right| \le \int_a^b |f(x)|\,dx\ .$$

When do we have equality?

8. Let $f : [a, b] \to \mathbb{R}$ be a continuous function. Then there exists $c \in (a, b)$ such that

$$\frac{1}{b-a}\int_a^b f(x)\,dx = f(c)$$

(see Theorem 50 above). Is this still true for Riemann integrable functions?

9. Consider the function

$$f(x) = \begin{cases} 1 & 0 \le x \le 1\ , \\ 0 & 1 < x \le 2\ . \end{cases}$$

(a) What is $F(x) = \int_0^x f(t)\,dt$ on $[0,2]$?

(b) Is $F(x)$ continuous?

(c) Is $F'(x) = f(x)$?

10. Let $f : [a,b] \to \mathbb{R}$ be a Riemann integrable function. Let $g : [a,b] \to \mathbb{R}$ be a function such that $\{x \in [a,b]; f(x) \neq g(x)\}$ is finite. Show that $g(x)$ is Riemann integrable and

$$\int_a^b f(x)\,dx = \int_a^b g(x)\,dx\ .$$

Does the conclusion still hold when $\{x \in [a,b]; f(x) \neq g(x)\}$ is countable?

11. Suppose that $f$ is Lipschitz with constant $L$ on $[a,b]$. Prove that

$$\left| \int_0^1 f(x)\,dx - \frac{1}{n}\sum_{i=1}^{n} f(\frac{i}{n}) \right| \leq \frac{L}{n}.$$

12. We say $f$ is Riemann integrable on $[a,\infty)$ if $\lim_{b\to\infty} \int_a^b f(x)\,dx$ exists. Use $\int_a^\infty f(x)\,dx = \lim_{b\to\infty} \int_a^b f(x)\,dx$, to show that $\int_0^\infty \frac{\sin x}{x}\,dx$ exists.

# Chapter 2

# Real Analysis

> The following theorem can be found in the work of Mr. Cauchy: "If the various terms of the series $u_1 + u_2 + \cdots$ are continuous functions, then the sum $s$ of the series is also a continuous function of $x$." But it seems to me that this theorem admits exceptions. For example the series
>
> $$\sin x - \frac{1}{2}\sin 2x + \frac{1}{3}\sin 3x...$$
>
> is discontinuous at each value $(2m+1)\pi$ of $x$, ...
>
> *Abel, 1826, Oeuvres, vol. 1, p. 224–225*

## 2.1 Metric, Normed, and Inner Product Spaces

Think about all of our usual operations, starting from limits, derivatives, integrals, sums, then all operations on sets and functions, vectors, matrices, etc. Instead of dealing with individual operations, M. Fréchet, together with other mathematicians of the twentieth century, developed abstract spaces, where these operations were considered as functions defined on entire collections of abstract objects. As early as 1906, in his thesis, "Sur quelques points du calcul fuctionnel," Fréchet introduced a notion of distance defined on abstract sets of points (Figure 2.1). He considered the collection $C[0,1]$ of continuous functions $f : [0,1] \to \mathbb{R}$ where one measures the distance between functions $d(f,g)$ by taking the maximum vertical distance between their graphs, that is:

$$d(f,g) = \max_{0 \le x \le 1} |f(x) - g(x)|.$$

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

A. G. Aksoy, *Fundamentals of Real and Complex Analysis*, Springer Undergraduate Mathematics Series, https://doi.org/10.1007/978-3-031-54831-4_2

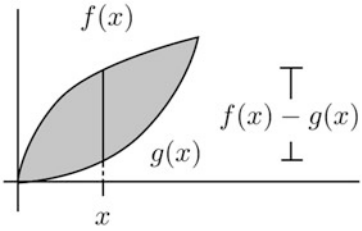



Figure 2.1: Max vertical distance

In 1928, Fréchet also wrote a monograph on his research of abstract spaces called "Les Espaces Abstraits." This monograph contains most of the examples of abstract metric spaces. The other two spaces of interest are normed and inner product spaces. In 1922, S. Banach wrote down the axioms for normed spaces and axioms of inner product spaces, which were later presented by John von Neumann. In the following, we first define a *"metric space," "inner product space,"* and *"normed space."* We will see that every inner product space is a normed space and every normed space is a metric space. All of these abstract spaces are generally called *"topological spaces."*

## Metric Spaces

In this section we will introduce the notion of a metric space. The word *metric* is simply a synonym for *distance*.

**Definition 39** (Metric)**.** Given a nonempty set $M$, a function

$$d : M \times M \to \mathbb{R}$$

satisfying the following properties is called a *metric* on $M$.

a) $0 \leq d(x, y) < \infty$ for all pairs $x, y \in M$.

b) $d(x, y) = 0 \Leftrightarrow x = y$.

c) $d(x, y) = d(y, x)$ for all pairs $x, y \in M$.

d) $d(x, y) \leq d(x, z) + d(z, y)$ for all $x, y, z \in M$ (Triangle Inequality).

We can use this function to define the concept of a metric space.

**Definition 40** (Metric Space)**.** The couple $(M, d)$ consisting of a set together with a metric $d$ defined on $M$ is called a *metric space*.

There are many different kinds of metric spaces, some of which play very significant roles in geometry and analysis. We will now look at some common examples of metric spaces. Our first example is a rather trivial metric, but it shows every set $M$ admits at least one metric.

**Example 55.** Let $M$ be an arbitrary nonempty set. Define $d$ by

$$d(x,y) = \begin{cases} 0 & \text{if } x = y \\ 1 & \text{if } x \neq y \end{cases}.$$

This always available metric is called the *discrete metric* and is a useful metric to remember if one is looking for a counterexample.

**Example 56.** The usual absolute value defines a metric on $\mathbb{R}$:

$$d(x,y) = |x - y|.$$

Anytime we refer to $\mathbb{R}$ without explicitly naming a metric, the *absolute value metric* is always understood to be the one we have in mind. It is useful to note that by restricting the metric to a subset, any subset of a metric space can be considered as a metric space.

**Example 57.** Let $A$ be a subset of $M$; if we restrict $d(x,y)$ to $A$, we define $d(x,y)$ on $A \times A$ satisfying the properties $a)$ to $d)$ above using the same letter $d$ and say $(A,d)$ is a metric space. In particular, when we consider $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R} \setminus \mathbb{Q}$, each come already supplied with a natural metric, namely the restriction of the usual absolute value metric on $\mathbb{R}$.

**Example 58.** In $\mathbb{R}^n$ the notion of $n$-dimensional Euclidean distance produces a metric. If $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ are elements in $\mathbb{R}^n$, then

$$d(x,y) = \left((x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2\right)^{1/2}$$

defines a metric on $\mathbb{R}^n$. This is the *usual metric* on $\mathbb{R}^n$. In order to justify this claim we need to check the above conditions from a) to d). The triangle inequality is the one which requires attention and we will give a proof of it when we see "Minkowski's Inequality" later.

**Remark 31.** There is an important collection of metrics on $\mathbb{R}^n$. Let $p$ be a real number such that $p \geq 1$. For $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$, we can define:

$$d_p(x,y) = ||x - y||_p$$

where by $||x||_p$, we mean

$$||x||_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}.$$

To show $d_p$ is a metric on is a metric on $\mathbb{R}^n$, we need the following *Hölder's Inequality*. See, e.g., [22], pp. 182–196 for details.

*Hölder's Inequality*: Let $p$ and $q$ be positive numbers such that $\dfrac{1}{p} + \dfrac{1}{q} = 1$. Then for arbitrary real numbers $a_1, a_2, \cdots, a_n$ and $b_1, b_2, \cdots, b_n$,

$$\left|\sum_{k=1}^{n} a_k b_k\right| \leq \left(\sum_{k=1}^{n} |a_k|^p\right)^{\frac{1}{p}} \left(\sum_{k=1}^{n} |b_k|^q\right)^{\frac{1}{q}}.$$

Unless $a_k \equiv 0$ or $b_k \equiv 0$, equality occurs if and only if $a_k b_k$ has constant sign and $|a_k|^p = \lambda |b_k|^q$ for some number $\lambda > 0$ and all $k = 1, 2, \ldots, n$.
For $p = \infty$, the metric $d_p(x, y) = ||x - y||_p$ will be

$$d_\infty(x, y) = ||x - y||_\infty = \max_{1 \le i \le n} |x_i - y_i|.$$

It is straightforward to prove that $d_\infty$ is a metric on $\mathbb{R}^n$. The space $(\mathbb{R}^n, d_p)$ is sometimes denoted by $\ell_n^p(\mathbb{R})$.

**Example 59.** The usual metric is not the only metric on $\mathbb{R}^2$. Given $x = (x_1, x_2)$ and $y = (y_1, y_2)$ in $\mathbb{R}^2$, one can define:

$$d(x, y) = |x_1 - y_1| + |x_2 - y_2|.$$

This is sometimes called *taxicab metric*. This metric is illustrated in Figure 2.2.

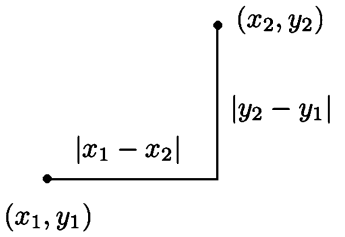



Figure 2.2: The taxicab metric

**Example 60.** Let $C[a, b]$ represent the collection of all continuous real-valued functions $f : [a, b] \to \mathbb{R}$. Then,

$$d(f, g) = ||f - g|| = \sup_{x \in [a,b]} |f(x) - g(x)|$$

is a metric on $C[a, b]$. It is clear that $d(f, g)$ is symmetric. Note that $f - g$ being a continuous function on the compact set $[a, b]$ means it is a bounded function by the Extreme Value Theorem. Moreover $d(f, g) \ge 0$ and $d(f, g) = 0$ if and only if $f(x) = g(x)$ for each $x \in [a, b]$. To show the triangle inequality observe:

$$d(f, g) = \sup_{x \in [a,b]} |f(x) - g(x)| = \sup_{x \in [a,b]} |f(x) - h(x) + h(x) - g(x)|$$

and since

$$\sup_{x \in [a,b]} |f(x) - h(x) + h(x) - g(x)| \le \sup_{x \in [a,b]} |f(x) - h(x)| + \sup_{x \in [a,b]} |h(x) - g(x)|,$$

we have $d(f, g) \le d(f, h) + d(h, g)$ for all $f, g$, and $h$ in $C[a, b]$.

We will see later that convergence in this metric space means uniform convergence (see Theorem 91 in Section 2.3). Furthermore recall the fact about "sup" of a sum of a two sets given in Remark 5 in Section 1.3.

## Topology of Metric Spaces

The fundamental and necessary notion in understanding metric spaces is an open ball.

**Definition 41.** Suppose that $(M, d)$ is a metric space and $a \in M$. If $r$ is a positive real number, the *open ball* centered at $a$ of radius $r$ is a subset of $M$ defined by

$$B_r(a) = \{x \in M : \ d(x, a) < r\}.$$

The *closed ball* centered at $a$ of radius $r$ is a subset of $M$ defined by

$$\overline{B_r(a)} = \{x \in M : \ d(x, a) \leq r\}.$$

Notice that in $\mathbb{R}$, open balls are open intervals $B_r(a) = (a - r, a + r)$ and closed balls are closed intervals $\overline{B_r(a)} = [a-r, a+r]$. If $(M, d)$ is a metric space with the discrete metric, then $B_r(a) = \overline{B_r(a)} = \{a\}$ for all $0 < r < 1$.

The following figure illustrates the unit balls (i.e., when $r = 1$) centered at $(0, 0)$ in $\mathbb{R}^2$ with respect to several metrics (Figure 2.3):

a) $d(x, y) = \left((x_1 - y_1)^2 + (x_2 - y_2)^2\right)^{1/2}$ (the usual metric).

b) $d_1(x, y) = |x_1 - y_1| + |x_2 - y_2|$ (taxicab metric).

c) $d_\infty(x, y) = \max\{|x_1 - y_1|, |x_2 - y_2|\}$ (uniform metric).

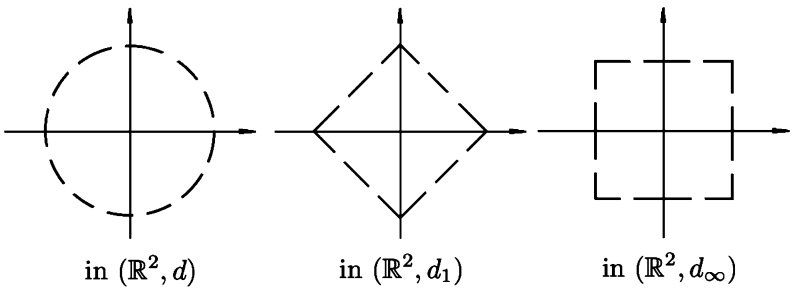



Figure 2.3: The unit balls in $\mathbb{R}^2$ with respect to various metrics

**Definition 42.** Let $(M, d)$ be a metric space and $A \subset M$. The set $A$ is an *open* set in $M$ if, for each $a \in A$ there is an $r > 0$ such that $B_r(a) \subseteq A$. A set $F \subseteq M$ is called closed if $F^c := M \setminus F$ is open (Figure 2.4).

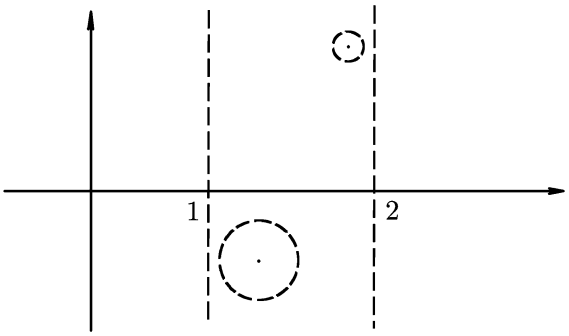



Figure 2.4: Radius $r$ of $B_r(a)$ depends on the point $a$

**Proposition 18.** *An open ball in any metric space $(M, d)$ is an open set. Every closed ball is closed set (Figure 2.5).*

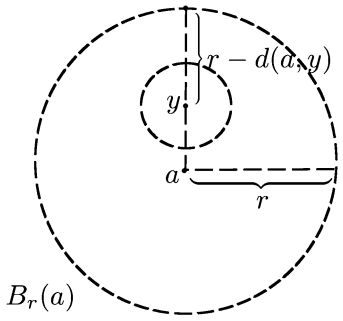



Figure 2.5: An open ball is an open set

*Proof.* Let $B_r(a) = \{x \in M : d(x,a) < r\}$ be an open ball with radius $r > 0$. For any $y \in B_r(a)$ want to find an $\epsilon > 0$ such that $B_\epsilon(y) \subset B_r(a)$. Let $\epsilon = r - d(a,y)$; since $y \in B_r(a)$, we have $\epsilon > 0$. Let $z \in B_\epsilon(y)$, then by the triangle inequality we have

$$d(a,z) \leq d(a,y) + d(y,z) < d(a,y) + \epsilon = r$$

which implies $z \in B_r(a)$ and $B_\epsilon(y) \subset B_r(a)$ as desired. Similarly, one can show that $\{x \in M : \ d(x,a) > r\}$ is an open set. Hence the closed ball is closed. □

**Proposition 19.** *In any metric space $M$, the empty space $\emptyset$ and the full space $M$ are open sets.*

*Proof.* To show the empty set $\emptyset$ is open, we must show that each point in $\emptyset$ is the center of an open ball contained in $\emptyset$. Since there are no points in $\emptyset$, this requirement is automatically satisfied. On the other hand since every open ball centered on each point is clearly contained in $M$, this implies that $M$ is an open set. □

**Remark 32.** Note that being "open" and "closed" are related by complementation not with negation. If a set is not open, we cannot deduce it is closed or vice versa. Some sets are neither open nor closed, like the interval $(0, 1]$. Moreover, in an arbitrary metric space $M$, the empty set $\emptyset$ and the whole space $M$ are also both closed and for some metric spaces, such as $\mathbb{R}^n$, these are only two sets which are simultaneously open and closed. For other metric spaces, the situation could be different as shown in the following proposition.

**Proposition 20.** *Consider $(\mathbb{R}, d)$ where $d(x, y)$ is the discrete metric. Then every set is both open and closed.*

*Proof.* It is sufficient to prove that every subset of $\mathbb{R}$ is open with respect to the discrete metric. Let $A \neq \emptyset$ and $A \subseteq \mathbb{R}$. Let $a \in A$ be an arbitrary point; since $B_1(a) = \{a\} \subset A$, $A$ is open. □

The fundamental properties of open sets in a metric space $M$ are stated in the next theorem, which says that the class of all open sets in a metric space is closed under the formation of arbitrary unions and finite intersections.

**Theorem 54.** *Let $M$ be a metric space. Then*

*a) Any union of open sets in $M$ is open.*

*b) Any finite intersection of open sets in $M$ is open.*

*Proof.* Let $\{G_j\}_j$ be a family of open sets in a metric space $(M, d)$ and suppose that $x \in \bigcup_{j \in J} G_j$. Then $x \in G_k$ for some $k \in J$. Since $G_k$ is open, there is a real number $r > 0$ such that $B_r(x) \subset G_k$. Then $B_r(x) \subseteq \bigcup_{j \in J} G_j$ and this completes the proof of part a).

For part b), suppose $x \in \bigcap_{j=1}^{n} G_j$, then $x \in G_j$ for each $j = 1, 2, \ldots, n$. Since $G_j$ is open for each $j$, there exists a $r_j$ such that $B_{r_j}(x) \subseteq G_j$. Let $r = \min\{r_1, r_2, \ldots, r_n\}$. Then $r > 0$ and $B_r(x) \subseteq \bigcap_{j=1}^{n} G_j$. □

Even though the structure of the open sets can be complicated in an arbitrary metric space, in the case of the real line one has the following intuitive description of open sets.

**Theorem 55.** *Every nonempty open set in $\mathbb{R}$ is the union of a countable disjoint collection of open intervals.*

For the proof of the Theorem 55 we refer the reader to the proof of the Theorem 27.

## Convergence in a Metric Space

Next, we examine the convergence of sequences in arbitrary metric spaces. If you consider the similarity between a metric $d(x, y)$ and the distance expressed as the absolute value, we see that we can transfer much of the theory of limits of sequences from $\mathbb{R}$ to an arbitrary metric space.

**Definition 43.** Let $\{x_n\}$ be a sequence in a metric space $M$.

a) We say a sequence $\{x_n\}$ *converges* in $(M, d)$ if there is a point $x \in M$ such that for every $\epsilon > 0$ there is an $N \in \mathbb{N}$ such that

$$n \geq N \quad \text{implies} \quad d(x_n, x) < \epsilon.$$

In such case, we use the notation $\{x_n\} \to x$ or simply write $x_n \to x$.

b) We say $\{x_n\}$ is *Cauchy* if for every $\epsilon > 0$ there is an $N \in \mathbb{N}$ such that

$$n, m \geq N \quad \text{implies} \quad d(x_n, x_m) < \epsilon.$$

c) A sequence $\{x_n\}$ is said to be *bounded* if there is a constant $K > 0$ and a $y \in M$ such that

$$d(x_n, y) \leq K \quad \text{for all} \quad n \in \mathbb{N}.$$

**Theorem 56.** *Let $(M, d)$ be a metric space.*

*a) If a sequence $\{x_n\}$ converges in $(M, d)$ to a limit $x$, then $x$ is unique.*

*b) If a sequence $\{x_n\} \to x$ in $(M, d)$ and $\{x_{n_k}\}$ is a subsequence of $\{x_n\}$, then $x_{n_k} \to x$ (as $k \to \infty$) as well.*

*c) Every convergent sequence in $(M, d)$ is bounded.*

*d) Every convergent sequence in $(M, d)$ is a Cauchy sequence.*

*Proof.* Modify the proofs of the similar theorems given in Chapter 1, Section 1.4 replacing $|x - y|$ by the given metric $d(x, y)$. □

Now while several results about sequences in $\mathbb{R}$ carry over to an abstract metric space $(M, d)$, there are exceptions. For example, Cauchy sequences need not converge to a point in $M$. For example take $M = (0, 1]$ with the usual metric and the sequence $\{x_n\} = \left\{\frac{1}{n}\right\}_1^\infty \subset (0, 1]$. Then even though $\{x_n\}$ is Cauchy, it does not converge to any point in $M$.

## Completeness of Metric Spaces

**Definition 44.** A metric space $(M, d)$ is called *complete* if every Cauchy sequence of points of $M$ converges to a point of $M$.

Note that completeness requires two properties, one is every Cauchy sequence in $M$ is convergent and the other condition is that the limit of every Cauchy sequence stays in $M$. The Bolzano–Weierstrass theorem holds for $\mathbb{R}^n$ and thus $\mathbb{R}^n$ is complete for all $n \in \mathbb{N}$ (you prove it). On the other hand $(\mathbb{Q}, d)$ with $d(x, y) = |x - y|$ is a metric space. Recalling the density of the rationals, we can choose a sequence $\{x_n\}$ in $(\mathbb{Q}, d)$, such that $x_n \to \sqrt{2}$. Take

$$x_1 = 1.4,\ x_2 = 1.41,\ x_3 = 1.414, \ldots, x_n, \ldots$$

where

$$x_n = (n+1)\text{-digits in the decimal expansion of } \sqrt{2}.$$

Then the sequence is Cauchy (why?), but the limit point $\sqrt{2} \notin \mathbb{Q}$. Thus $(\mathbb{Q}, d)$ is not a complete metric space.

**Remark 33.** Cauchy sequences always converge in a discrete space. Because in a discrete metric space, a sequence $\{x_n\}$ is Cauchy if and only if it is eventually constant, that is for sufficiently large $N$ we have $x_n = x$ for some fixed $x$.

In the following we give two additional examples of metric spaces which are not complete.

**Example 61.**

a) Let $\mathcal{P}$ be the set of all polynomials (of all degrees) defined on $[0, 1]$. Define $d(x, y)$ by

$$d(x, y) = \max_{0 \le t \le 1} |x(t) - y(t)|.$$

Then, $(\mathcal{P}, d)$ is not a complete metric space. It is clear that $(\mathcal{P}, d)$ is a metric space. To show $(\mathcal{P}, d)$ is not complete, consider the following sequence:

$$x_n(t) = \sum_{k=0}^{n} \left(\frac{t}{2}\right)^k = 1 + \frac{t}{2} + \cdots + \frac{t^n}{2^n} \quad 0 \le t \le 1.$$

Clearly, $x_n(t) \in \mathcal{P}$ for each $n \in \mathbb{N}$. Next, we show that the sequence $\{x_n\}$ is a Cauchy sequence. Taking $m < n$ we observe

$$\begin{aligned} d(x_n, x_m) &= \max_{0 \le t \le 1} |x_n(t) - x_m(t)| \\ &= \max_{0 \le t \le 1} \left| \sum_{k=0}^{n} \left(\frac{t}{2}\right)^k - \sum_{k=0}^{m} \left(\frac{t}{2}\right)^k \right| \\ &= \max_{0 \le t \le 1} \left| \sum_{k=m+1}^{n} \left(\frac{t}{2}\right)^k \right| \end{aligned}$$

$$\begin{aligned} &\leq \max_{0 \leq t \leq 1} \left| \sum_{k=m+1}^{n} \frac{1}{2^k} \right| \\ &= \frac{1}{2^m} - \frac{1}{2^n}. \end{aligned}$$

This difference is arbitrarily small for large enough $m$ and $n$, which implies that $\{x_n\}$ is a Cauchy sequence in $\mathcal{P}$. However, this sequence does not converge in $(\mathcal{P}, d)$, because the only candidate for $\lim_{n\to\infty} x_n(t) = \dfrac{2}{2-t}$, for $0 \leq t \leq 1$, and this is not a polynomial function. Since not every Cauchy sequence converges in $(\mathcal{P}, d)$, we conclude that $(\mathcal{P}, d)$ is not complete.

b) While $C[0,1]$ with the metric

$$d(f,g) = \sup\{|f(x) - g(x)| : \quad x \in [0,1]\}$$

is a complete metric space, $C[0,1]$ is not a complete metric space if we complete metric space if we take the metric

$$d(f,g) = \int_0^1 |f(x) - g(x)|\, dx$$

on it. For details we refer the reader to, e.g.,

Let $(M,d)$ be a metric space and $S$ a nonempty subset of $M$. As we defined earlier, by the restriction of $d$ to $S$ we mean the mapping

$$d_S : S \times S \to [0,\infty) \quad \text{such that} \quad d_S(x,y) = d(x,y)$$

for $x, y \in S$. It is clear that $d_S$ is a metric on $S$ and we usually drop $S$ and express this metric on $S$ as $d$ as well. The metric space $(S,d)$ is called a subspace of $(M,d)$ and a subspace $(S,d)$ of the metric space $(M,d)$ is said to be *closed* if it contains the limits of all sequences in $S$ which converge in $M$. The next theorem guarantees the completeness of many metric spaces which arise as subspaces of complete metric spaces.

**Theorem 57.** *Let $(M,d)$ be a complete metric space and let $(S,d)$ be a subspace of $M$. Then $S$ is complete if and only if it is closed.*

*Proof.* Assume that $S$ is closed; we want to show that it is complete. Let $\{x_n\}$ be a Cauchy sequence in $S$. It is also a Cauchy sequence in $(M,d)$, and since $M$ is complete $\{x_n\}$ converges to a point $x$ in $M$; but since $S$ is closed, the limit of the sequence must belong to $S$, so $S$ is complete. Now let us prove the converse. Suppose $S$ is complete, let $\{x_n\}$ be any sequence in $S$ and suppose that as a sequence in $M$ it converges with limit $x$. Then $\{x_n\}$ is a Cauchy sequence in $M$ and hence also in $S$. Since $S$ is complete, we must have $x \in S$, so $S$ is closed. □

Now we know that in a complete metric space, the notions of completeness and closedness for subspaces coincide. Another natural question in regards to the completeness of metric spaces is whether or not an arbitrary incomplete metric space can be somehow enlarged to a metric space which is complete. After all, $\mathbb{Q}$ is an incomplete metric space but we can enlarge to the real line $\mathbb{R}$ which is complete. This question has a positive answer, but we first need the following definition:

**Definition 45.** Let $(M, d)$ and $(\tilde{M}, \tilde{d})$ be metric spaces. A mapping $T : M \to \tilde{M}$ is said to be *isometric* if $T$ preserves distances. That is for all $x, y \in M$,

$$\tilde{d}(Tx, Ty) = d(x, y).$$

The space $M$ is said to be *isometric* with the space $\tilde{M}$ if there is a bijective (one-to-one and onto) isometry of $M$ onto $\tilde{M}$. In this case the spaces $M$ and $\tilde{M}$ are called *isometric spaces*.

We can now state the following theorem without proof (for proof consult, e.g., [35], pp. 42–45):

**Theorem 58.** *For a metric space $(M, d)$ there is a complete metric space $(\tilde{M}, \tilde{d})$ (called the completion) which has the following properties:*

*a) $(\tilde{M}, \tilde{d})$ has a subspace $S$ that is isometric to $(M, d)$.*

*b) $(M, d)$ is dense in $(\tilde{M}, \tilde{d})$.*

*c) $(\tilde{M}, \tilde{d})$ is unique except for isometries (i.e., if $\hat{M}$ is any complete metric space having a dense subspace $\hat{S}$ isometric with $M$, then $\hat{M}$ and $\tilde{M}$ are isometric as well).*

## Completion and Ultrametric

The set of real numbers $\mathbb{R}$ is the completion of $\mathbb{Q}$ with respect to Euclidean norm where the induced metric by this norm is $d(x, y) = |x - y|$. *The field of p-adic numbers* $\mathbb{Q}_p$ defined for each prime $p$ is constructed from the rational numbers $\mathbb{Q}$ as its completion with respect to a certain norm which depends on the prime number $p$. This norm is quite different from the Euclidean norm.

**Definition 46.** Let $F$ be a field. A *norm* on $F$ is a map denoted by

$$||\cdot|| : F \to \mathbb{R} \geq 0$$

satisfying the properties:

a) $||x|| = 0$ if and only if $x = 0$.

b) $||x\, y|| = ||x||||y||$ for all $x, y \in F$.

c) $||x + y|| \leq ||x|| + ||y||$ for all $x, y \in F$ (The Triangle Inequality).

This norm induces a metric (or a distance function) by $d(x, y) = ||x - y||$ and we can regard $(F, d)$ as a metric space.

**Definition 47.** Let $||\cdot||_1$ and $||\cdot||_2$ be two norms on a field $F$. We say that these two norms are *equivalent* if and only if there exists a positive real number $\alpha$ such that $||x||_2 = ||x||_1^\alpha$ for every $x \in F$.

This means that a sequence is Cauchy with respect to $||\cdot||_1$ if and only if it is Cauchy with respect to $||\cdot||_2$. We have already seen that the absolute value $|\cdot|$ is a norm on $\mathbb{Q}$. The proposition below describes all the norms on $\mathbb{Q}$ which are equivalent to the absolute value norm.

**Proposition 21.** *Let $\alpha > 0$. $||x|| = |x|^\alpha$ is a norm on $\mathbb{Q}$ if and only if $\alpha \leq 1$. In this case it is equivalent to the absolute value norm.*

*Proof.* First note that if $\alpha > 1$ then the triangle inequality fails since $|1+1|^\alpha = 2^\alpha > |1|^\alpha + |1|^\alpha = 2$. Suppose $\alpha \leq 1$. The first two properties of the norm are straightforward, so we need to check the triangle inequality. To show the triangle inequality holds, we first observe the behavior of $t^\alpha$. Note that for $t \geq 1$, $t^\alpha \leq t$ and for $0 \leq t \leq 1$ we have $t^\alpha \geq t$. Assume $\dfrac{|y|}{|x|} \leq 1$. Then by setting $t = 1 + \dfrac{|y|}{|x|}$, we have:

$$|x+y|^\alpha \leq (|x| + |y|)^\alpha = |x|^\alpha \left(1 + \frac{|y|}{|x|}\right)^\alpha \leq |x|^\alpha \left(1 + \frac{|y|}{|x|}\right) \leq |x|^\alpha \left(1 + \frac{|y|^\alpha}{|x|^\alpha}\right)$$

and therefore

$$|x+y|^\alpha \leq |x|^\alpha + |y|^\alpha. \qquad \square$$

There is a theorem due to Ostrowski [30], p. 43 which describes in a precise way all norms on $\mathbb{Q}$ which are equivalent to the absolute value $|\cdot|$. We will see this after we explain the $p$-adic absolute value $|\cdot|_p$ on $\mathbb{Q}$.

**Definition 48.** A norm is called *non-Archimedian* if it satisfies the non-Archimedian triangle inequality:

$$||x + y|| \leq \max\{||x||, ||y||\}.$$

The metric induced by a non-Archimedian norm is said to be an *ultrametric*. Note that instead of the triangle inequality for the usual metric

$$d(x, z) \leq d(x, y) + d(y, z),$$

an ultrametric satisfies the stronger triangle inequality:

$$d(x, z) \leq \max\{d(x, y), d(y, z)\}.$$

Metric spaces with ultrametrics are called *ultrametric spaces*.

**Example 62.** We have already seen one example of an ultrametric space. The discrete metric is an ultrametric (why?). Perhaps the most important examples of ultrametric spaces come from the $p$-adic numbers; $p$-adic numbers form a complete ultrametric space. Before we define *$p$-adic number*s and *$p$-adic absolute value*, let us make one nice geometrical observation about ultrametric spaces.

**Proposition 22.** *Any triangle in an ultrametric space is isosceles and the length of its base does not exceed the length of sides.*

*Proof.* Suppose $a$ and $x$ are elements of a non-Archimedian field $F$, satisfying $||x-a|| < ||a||$, then using the non-Archimedian property

$$||x|| = ||x - a + a|| \leq \max\{||x-a||, ||a||\} = ||a||.$$

On the other hand

$$||a|| = ||a - x + x|| \leq \max\{||x-a||, ||x||\}.$$

Now if $||x-a|| > ||a||$, then by the above inequality we get $||a|| \leq ||x-a||$ contrary to the assumption, thus

$$||a|| = ||a - x + x|| \leq \max\{||x-a||, ||x||\} \leq ||x||.$$

So, $||x|| = ||a||$. What we showed can be restated as follows: let $a, b$ be elements in an ultrametric space, then

$$||b|| < ||a|| \quad \text{implies} \quad ||a+b|| = ||a||.$$

In other words, the longer side "wins" or if $a, b$, and $a+b$ are the sides of a triangle in an ultrametric space then this triangle is isosceles and the length of its base does not exceed the length of the sides. □

## $p$-adic Absolute Value Metric

As mentioned above on the field $\mathbb{Q}$ we have the Euclidean metric to measure the distance between two rational numbers. It is natural to wonder if the Euclidean distance is the most useful one. Is there any other way one can describe the closeness between rational numbers? It turns out that the $p$-adic absolute value function measures distance between rational numbers from the following construction.

**Definition 49.** Let $p \in \mathbb{Z}$ be a prime number. For any nonzero integer $a$ *$p$-adic ordinal* of $a$ is denoted by $\text{ord}(a)$ and defined as

$$\text{ord}_p(a) = \text{the largest power of } p \text{ which divides } a.$$

If $x = \dfrac{a}{b}$ where $a, b \in \mathbb{Z}$, $b \neq 0$ then

$$\text{ord}_p(x) = \text{ord}_p(a) - \text{ord}_p(b).$$

**Example 63.** The following are simple examples for the concept of $p$-adic ordinals.

- $\text{ord}_5(22) = 0$ since $5 \nmid 22$.
- $\text{ord}_3(423) = 2$ since $423 = 3^2 \cdot 47$ and $3 \nmid 47$.
- $\text{ord}_2\left(\frac{4}{7}\right) = \text{ord}_2(2^2) - \text{ord}_2(7) = 2 - 0 = 2$.

**Definition 50.** Let $p \in \mathbb{Z}$ be a prime number. For any rational number $x$, define a map $|\cdot|_p$ on $\mathbb{Q}$ called the *p-adic absolute value* as follows:

$$|x|_p = \begin{cases} p^{-\text{ord}_p(x)} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

Going back to the above example $|22|_5 = 5^0 = 1$, $|423|_3 = 3^{-2} = \frac{1}{9}$ and $\left(\frac{4}{7}\right)|_2 = 2^{-2} = \frac{1}{4}$.

**Proposition 23.** *$|\cdot|_p$ is a non-Archimedian norm on $\mathbb{Q}$. The p-adic absolute value $|\cdot|_p$ gives a metric on $\mathbb{Q}$ defined by*

$$d(x, y) = |x - y|_p \quad x, y \in \mathbb{Q}.$$

*Proof.* We need to show

a) $|x|_p \geq 0$ and $|x|_p = 0$ if and only if $x = 0$.

b) $|xy|_p = |x|_p|y|_p$.

c) $|x + y|_p \leq \max\{|x|_p, |y|_p\}$.

a) is clear from the definition and b) follows from

$$\text{ord}_p(xy) = \text{ord}_p(x) + \text{ord}_p(y).$$

It remains to show part c) holds. Note that if $x = 0$ or $y = 0$, then c) is true, so without loss of generality assume $x, y \neq 0$. Let $x = \frac{a}{b}$ and $y = \frac{c}{d}$, then we have $x + y = \frac{ad + bc}{bd}$ and

$$\text{ord}_p(x + y) = \text{ord}_p\left(\frac{ad + bc}{bd}\right) = \text{ord}_p(ad + bc) - \text{ord}_p(bd).$$

Now observe that the highest power of $p$ dividing the sum $ad + bc$ is greater than or equal to both the power of $p$ that divides $ad$ and the power of $p$ that divides $bc$. Therefore,

$$\text{ord}_p(ad + bc) \geq \min\{\text{ord}_p(ad), \text{ord}_p(bc)\}.$$

Furthermore, the highest power $p$ that divides the product $bd$ is just the sum of the power of $p$ dividing $b$ and dividing $d$, i.e., $\text{ord}_p(bd) = \text{ord}_p(b) + \text{ord}_p(d)$. Combining this information we obtain

$$\text{ord}_p(x+y) \geq \min\{\text{ord}_p(ad), \text{ord}_p(bc)\} - \text{ord}_p(b) - \text{ord}_p(d)$$

or equivalently

$$\text{ord}_p(x+y) \geq \min\{\text{ord}_p(x), \text{ord}_p(y)\}.$$

It follows that $-\text{ord}_p(x+y) \geq \max\{\text{ord}_p(x), \text{ord}_p(y)\}$. Therefore,

$$|x+y|_p = p^{-\text{ord}_p(x+y)} \leq \max\{p^{-\text{ord}_p(x)}, p^{-\text{ord}_p(y)}\} = \max\{|x|_p, |y|_p\}.$$

Note that $|\cdot|_p$ also satisfies the strong triangle inequality

$$|x+y|_p \leq \max\{|x|_p, |y|_p\} \leq |x|_p + |y|_p. \qquad \square$$

The $p$-adic metric plays a significant role in analysis as well as in number theory. $\mathbb{Q}$ admits the $p$-adic norm $|\cdot|_p$ for each prime $p$ as well as the absolute value $|\cdot|$. The following theorem, due to Ostrowski, states that this list is complete in some sense.

**Theorem 59** (Ostrowski's Theorem)**.** *Every nontrivial norm* $||\cdot||$ *on* $\mathbb{Q}$ *is either equivalent to* $|\cdot|_p$ *for some* $p$ *or equivalent to* $|\cdot|$.

For the proof of Ostrowski's theorem we refer to, e.g., [30], pp. 43–45. One can now define *Cauchy sequences with respect to* $|\cdot|_p$ in the usual way.A sequence of rational numbers $\{r_i\}$ is Cauchy in $|\cdot|_p$ if for every $\epsilon > 0$, there is an integer $N$ such that

$$|r_i - r_j|_p < \epsilon \quad \text{whenever } i, j > N.$$

It turns out that there is a Cauchy sequence in $\mathbb{Q}$ with respect to $|\cdot|_p$ that does not converge in $\mathbb{Q}$ (consult [30], pp. 43–45 for a detailed discussion or for an example, set $x_n = 1 + p + \cdots + p^n$, then show $\{x_n\}$ is Cauchy in $(\mathbb{Q}, |\cdot|_p)$ and does not converge to any $x \in \mathbb{Q}$). This shows that $\mathbb{Q}$ is not complete with respect to $|\cdot|_p$.

**Definition 51.** Let $p$ be a fixed prime. *The field of p-adic numbers* is denoted by $\mathbb{Q}_p$ and defined to be the completion of $\mathbb{Q}$ with respect to $p$-adic norm $|\cdot|_p$.

Completing $\mathbb{Q}$ to the real numbers $\mathbb{R}$ is carried out by Cauchy sequences or by Dedekind cuts; similarly we think of elements of $\mathbb{Q}_p$ as equivalence classes of Cauchy sequences from $\mathbb{Q}$ with respect to $|\cdot|_p$. We can define addition and multiplication on $\mathbb{Q}_p$ so that it becomes a field. To learn more about algebraic properties of $\mathbb{Q}_p$, one can consult, e.g., [43], pp. 176–184.

## The Topology of $\mathbb{R}$ vs. the Topology of $\mathbb{Q}_p$

In this section we compare $\mathbb{R}$ and $\mathbb{Q}_p$ from the point of view of metric spaces. In $\mathbb{R}$ an open ball $B_r(a)$ with radius $r$ centered at $a$ is defined to be

$$B_r(a) := \{x \in \mathbb{R} : \ |x - a| < r\}.$$

In $\mathbb{Q}_p$ open balls are the sets

$$B_r(a)_p := \{x \in \mathbb{Q}_p : \ |x - a|_p < r\}.$$

However, since the $p$-adic norm assumes only discrete values $\{p^n | \ n \in \mathbb{Z}\} \cup \{0\}$, we only need to consider balls of radii $r = p^n$, where $n \in \mathbb{Z}$. The *sphere* in $\mathbb{Q}_p$ is defined by:

$$S_r(a)_p := \{x \in \mathbb{Q}_p : \ |x - a|_p = r\}.$$

There are several unexpected results that follow from the non-Archimedian property of $|\cdot|_p$. In the following we give a sample of such results.

**Proposition 24.** *Every triangle in $\mathbb{Q}_p$ is an isosceles triangle. That is given $x, y, z \in \mathbb{Q}_p$ at least two of the following distances are equal:*

$$|x - y|_p, \quad |y - z|_p, \quad |z - x|_p.$$

*Proof.* We have already shown this result for ultrametric spaces and $\mathbb{Q}_p$ is an ultrametric space. □

**Proposition 25.** *A sphere $S_r(a)_p$ is an open set in $\mathbb{Q}_p$.*

This is in sharp contrast with the situation in $\mathbb{R}$ or $\mathbb{R}^n$ where spheres are certainly not open sets.

*Proof.* To show $S_r(a)_p$ is an open set, for any $x \in S_r(a)_p$ and $\epsilon < r$ we must show $B_\epsilon(x)_p \subset S_r(a)_p$. Let $y \in B_\epsilon(x)_p$. Then

$$|x - y|_p < |x - a|_p = r.$$

Then by the above proposition

$$|y - a|_p = |x - a|_p = r,$$

and thus $y \in S_r(a)_p$. □

**Proposition 26.** *Every point in a ball is a center of that ball. In other words if $b \in B_r(a)_p$ then $B_r(a)_p = B_r(b)_p$.*

*Proof.* Let $x \in B_r(b)_p$, then $|x - b|_p < r$, and by hypothesis $|a - b|_p < r$. Therefore

$$|a - x|_p = |(a - b) + (b - x)|_p \leq \max\{|a - b|_p, |b - x|_p\} < r;$$

hence $B_r(b)_p \subset B_r(a)_p$. It is easy to see the other inclusion $B_r(a)_p \subset B_r(b)_p$. □

**Proposition 27.** *Two balls in $\mathbb{Q}_p$ have a nonempty intersection if and only if one is contained in the other.*

*Proof.* We want to show $B_r(a)_p \cap B_s(b)_p \neq \emptyset$ if and only if $B_s(b)_p \subset B_r(a)_p$ or $B_r(a)_p \subset B_s(b)_p$. Clearly if one ball is contained in the other then they have a nonempty intersection. Now suppose $r \leq s$ and $z \in B_r(a)_p \cap B_s(b)_p$. Since every point in a ball is its center, we have

$$B_r(a)_p = B_r(z)_p \quad \text{and} \quad B_s(b)_p = B_s(z)_p.$$

But $B_r(z)_p \subset B_s(z)_p$, and the required inclusions are satisfied. □

**Remark 34.** It is well known that the open balls in $\mathbb{Q}_p$ are both open and closed. Furthermore, the field $\mathbb{Q}_p$ has the Bolzano–Weierstrass property. That is if $A$ is a bounded infinite subset of $\mathbb{Q}_p$, then $A$ has an accumulation point in $\mathbb{Q}_p$. Additionally, $\mathbb{Q}_p$ also has the Heine–Borel property: if $A$ is a closed and bounded subset of $\mathbb{Q}_p$, then $A$ is compact. Note that $\mathbb{Q}_p$ is in many ways analogous to the field $\mathbb{R}$: both are complete metric spaces, both contain $\mathbb{Q}$ as a dense subset, but (as illustrated in the above propositions) the ultrametric on $\mathbb{Q}_p$ produces different properties for the topology of $\mathbb{Q}_p$.

## Compact Sets in Metric Spaces

**Definition 52.** A subset $S$ in a metric space $M$ is called *sequentially compact* if every sequence in $S$ has a subsequence that converges to a point in $S$.

Let $M$ be a metric space and $S \subset M$ a subset. An *open cover* for $S$ is a collection of open sets $\{U_i\}$ whose union also contains $S$; it is a *finite cover* if the collection contains only a finite number of sets.

**Definition 53.** A subset $S$ in a metric space is called *compact* if **every** open cover for $S$ has a finite subcover.

Note that in the above definition if one sets $S = M$, then one can talk about *compact metric spaces*. In the following we list some important theorems, without proof, about compactness in a metric space. Any classical analysis book contains most of these proofs. For example see, e.g., [26], pp. 151–160. The link that connects compactness to sequential compactness is the following theorem.

**Theorem 60** (Bolzano–Weierstrass Theorem)**.** *A subset of a metric space is compact if and only if it is sequentially compact.*

It is relatively easy to see a sequentially compact set is both closed and bounded. To see that it is closed, let $S$ be sequentially compact and take a sequence $\{x_n\} \in S$ converging to $x \in M$. Then since $S$ is sequentially compact, there is a subsequence of $\{x_n\}$ converging to some $s \in S$. Uniqueness of the limit implies that $s = x$, thus $S$ is closed. A sequentially compact set must also be bounded. If $S$ is sequentially compact and not bounded, then there is a sequence $\{x_n\} \in S$ and a point $s \in S$ with $d(x_n, s) \geq n$. Then $x_n$ cannot have any convergent subsequence.

**Example 64.** In Euclidean space we can easily identify a compact set because the Heine–Borel theorem states that a set $A \subset \mathbb{R}^n$ is compact if and only if it is closed and bounded. One half of this theorem is also true for a metric space. In fact a compact set is closed and bounded in a metric space too. But a set being closed and bounded does not imply compactness in an arbitrary metric space $M$. Here is an example: let $M$ be an infinite set with the discrete metric $d$ defined by

$$d(x,y) = \begin{cases} 0 & \text{if } x = y \\ 1 & \text{if } x \neq y \end{cases} .$$

The set $M$ is bounded because if $x_0 \in M$ be any point, then $M \subset B_2(x_0)$, where the open set $B_2(x_0) = \{y \in M |\ d(x_0, y) < 2\}$. Since M is the entire metric space it is also closed. However $M$ is not compact. Indeed, for each $x \in M$ we can form an open cover $\{B_{\frac{1}{2}}(x)\}_{x \in M}$ for $M$, but this open cover has no finite subcover, since we cannot delete any member $\{B_{\frac{1}{2}}(x)\}_{x \in M}$ while maintaining the property of being a cover of $M$.

Thus it is natural to ask if there is another characterization of compactness which exists in metric spaces. The answer is yes.

**Theorem 61.** *A metric space is compact if and only if it is complete and totally bounded.*

For proof and definition of totally bounded we again refer the reader to [26], p. 166. There is an important consequence of the Bolzano–Weierstrass theorem for metric spaces, called the Nested Sets Property.

**Theorem 62.** *Let $K_i$ be a sequence of compact nonempty sets in a metric space $M$ such that $K_{i+1} \subset K_i$ for all $i = 1, 2, \ldots$. Then there is at least one point in $\bigcap_{i=1}^{\infty} K_i$.*

This theorem is a consequence of the *finite intersection property* of compact sets, which we have given a version for the compact subsets of $\mathbb{R}$ in Section 1.5 of Chapter 1.

**Remark 35.** To describe the compact subsets of $C[0,1]$ we need the concept of equicontinuous family of functions. A sequence of functions $\{f_n\}$ defined on a set $A \subset \mathbb{R}$ is called *equicontinuous* if for every $\epsilon > 0$ there exists a $\delta > 0$ such that $|f_n(x) - f_n(y)| < \epsilon$ for all $n \in \mathbb{N}$ and $|x - y| < \delta$ in $A$. The Arzela–Ascoli theorem states that any bounded equicontinuous collection of functions in $C[0,1]$ must have a uniformly convergent subsequence. The proof of this theorem uses a clever diagonal selection process. See [26], p. 299–301.

## Baire's Category Theorem

Baire's Category Theorem is often regarded as one of the corner stones of functional analysis. The proofs of basic theorems in functional analysis, such as the

uniform boundedness principle and open mapping theorem, require completeness and depend on Baire's Category Theorem. There is also an unexpected consequence of this theorem: most continuous functions do not have derivatives at any point. Continuous functions that are differentiable at even one point in $[0, 1]$ form a "sparse" subset. We first state the theorem for a complete metric space.

**Theorem 63** (Baire's Theorem for a complete metric space)**.** *Let $M$ be a complete metric space. If $M = \bigcup_{n=1}^{\infty} E_n$, then the closure of some $E_n$ contains an open ball. Equivalently, if $(G_n)$ is a sequence of dense open sets in $M$, then $\bigcap_{n=1}^{\infty} G_n \neq \emptyset$; in fact $\bigcap_{n=1}^{\infty} G_n$ is dense in $M$.*

For the proof of the above theorem we refer the reader to, e.g., [14], p. 133.

In the following we state and give a proof of Baire's theorem for $\mathbb{R}$. There are many statements about the size of $\mathbb{R}$: for example since $\mathbb{R}$ is uncountable, we cannot write $\mathbb{R}$ as a countable union of single points. Baire's Theorem is an improved statement about the size of $\mathbb{R}$ again, which states that it is impossible to write the real numbers $\mathbb{R}$ as the countable union of nowhere-dense sets. The main ingredient to the proof of Baire's Theorem is the completeness of $\mathbb{R}$ which is given below. To understand what we mean by *nowhere-dense set*, and the motivation for defining such sets, we need to go back and see the formation of various subsets of $\mathbb{R}$. Every open subset of $\mathbb{R}$ is either a finite or countable union of open intervals, thus the structure of open sets in $\mathbb{R}$ is well understood. On the other hand if one thinks about the Cantor set, which is a closed subset of $\mathbb{R}$, it is uncountable and contains no open intervals of any type. Therefore we cannot expect a nice description of the closed subsets of $\mathbb{R}$ as in the case of its open subsets. Recall also that an arbitrary intersection of closed sets is closed and an arbitrary union of open sets is open. Thus one way to enlarge the list of subsets of $\mathbb{R}$ is to take the union of closed sets and intersection of open sets, which leads us to introduce the following concepts:

**Definition 54.** A subset $S \subseteq \mathbb{R}$ is called a $G_\delta$ set if it can be written as a countable intersection of open sets and a set $K \subseteq \mathbb{R}$ is called $F_\sigma$ if it can be written as a countable union of closed sets.

Using De Morgan's laws, it is not difficult to see that $S \subseteq \mathbb{R}$ is a $G_\delta$ set if and only if its complement $S^c$ is an $F_\sigma$ set. It turns out that a closed interval $[a, b]$ is a $G_\delta$ set, and there are some sets which are both an $F_\sigma$ set and a $G_\delta$ set. The half-open interval $(a, b]$ is one such example. We leave it to the reader to prove that the set of rationals $\mathbb{Q}$ is a $F_\sigma$ set and thus the set of irrationals $\mathbb{I}$ forms a $G_\delta$ set. Baire's Theorem makes it possible to claim that $\mathbb{Q}$ is not a $G_\delta$ set.

Recall that the set $S \subseteq \mathbb{R}$ is *dense* in $\mathbb{R}$ if given any two real numbers $x < y$, it is possible to find a point $s \in S$ with $x < s < y$. We have seen that the set $\mathbb{Q}$ is dense in $\mathbb{R}$ and clearly the set of integers $\mathbb{Z}$ is not dense in $\mathbb{R}$. The following are equivalent ways to express a particular set $S$ is dense in $\mathbb{R}$:

a) $S \subseteq \mathbb{R}$ is dense in $\mathbb{R}$ if and only if every point in $\mathbb{R}$ is a limit point of $S$.

b) $S \subseteq \mathbb{R}$ is dense in $\mathbb{R}$ if and only if $\overline{S} = \mathbb{R}$.

We now state a theorem (see, e.g., [14], p. 131). This theorem is about the dense open subsets of $\mathbb{R}$, but its conclusion reminds us of the result we have seen about a nested sequence of compact sets having a nontrivial intersection.

**Theorem 64.** *If $\{O_n\}$ is a sequence of dense open sets in $\mathbb{R}$, then $\bigcap_{n=1}^{\infty} O_n$ is not empty.*

**Definition 55.** A set $S$ is *nowhere-dense* in $\mathbb{R}$ if $\overline{S}$ contains no nonempty open intervals. Equivalently, a set $S$ is *nowhere-dense* in $\mathbb{R}$ if $\overset{\circ}{\overline{S}} = \emptyset$ (its closure has empty interior).

Nowhere-dense sets are small in certain sense; for example, the Cantor set and $\mathbb{Z}$ both are nowhere-dense in $\mathbb{R}$. Also every single element set $\{x\}$ is nowhere-dense in $\mathbb{R}$. Note that a set $S$ is nowhere-dense in $\mathbb{R}$ if it is dense in no (open) interval at all, or equivalently every interval $I \subset \mathbb{R}$ contains another interval $I^*$ such that $I^* \cap S = \emptyset$ (check the equivalence).

**Theorem 65** (Baire's Theorem for $\mathbb{R}$)**.** *The set of real numbers $\mathbb{R}$ cannot be written as the countable union of nowhere-dense sets.*

*Proof.* Suppose to the contrary that $\mathbb{R} = \bigcup_{n=1}^{\infty} A_n$ where every set $A_n$ is nowhere-dense in $\mathbb{R}$. Let $I_0 \subset \mathbb{R}$ be a closed interval of radius 1. Since $A_1$ is nowhere-dense in $\mathbb{R}$, it is nowhere-dense in $I_0$; thus there is a closed interval $I_1$ of radius less than $1/2$ such that $I_1 \subset I_0$ and $I_1 \cap A_1 = \emptyset$. Now observe that $A_2$ is nowhere-dense in $I_1$, thus there is a closed interval $I_2$ of radius less than $1/3$ such that $I_2 \subset I_1$ and $I_2 \cap A_2 = \emptyset$. Continuing in this fashion we get a sequence of closed intervals $\{I_n\}$ with radii converging to zero such that

$$I_n \cap A_n = \emptyset \quad \text{for} \quad n = 1, 2, \ldots$$

By the Nested Intervals Theorem the intersection $\bigcap_{n=1}^{\infty} I_n$ contains a point $x \in \mathbb{R}$. However, by the above construction $x \notin \bigcup_{n=1}^{\infty} A_n$, so it follows that $\mathbb{R}$ cannot be written as $\bigcup_{n=1}^{\infty} A_n$ where each $A_n$ is nowhere-dense in $\mathbb{R}$. □

**Corollary 12.** *$\mathbb{Q}$ cannot be written as the countable intersection of open sets.*

*Proof.* Observe that a dense $G_\delta$ subset of $\mathbb{R}$ must be also uncountable. □

**Remark 36.** Notice how very different $\mathbb{Q}$ and $\mathbb{I} = \mathbb{R} \setminus \mathbb{Q}$ are. Irrationals are quite "thick" while rationals are "sparse." As we have stated before, replacing $\mathbb{R}$ with an arbitrary complete metric space $M$, one can prove Baire's Theorem in this more general setting which is also called Baire's Category Theorem. The word *category* is there because it creates two categories of size for subsets of a metric space. A set qualifies to be "first category" if it can be written as a countable union of nowhere-dense sets; sometimes the word "meager" is used instead of first category. These are thin subsets of a metric space. Baire's Category Theorem asserts that a complete metric space is not the union of a countable collection of nowhere-dense sets. Thus, if the given metric space is complete, it is necessarily of the "second category," i.e., any complete metric space must be too large to be the countable union of nowhere-dense sets. Baire's Category Theorem plays an important role in many parts of analysis, in particular, in measure theory and the study of bounded linear operators on Banach spaces (a Banach Space is a complete normed linear space). In the following we state a very interesting and unexpected consequence of Baire's Theorem.

**Theorem 66** (An application of Baire's Theorem)**.** *Given the complete metric space $(C[0,1], d)$ where $d(f,g) = \sup_{x\in[0,1]} |f(x) - g(x)|$, let $S \subset C[0,1]$ be defined as*

$$S = \{f \in C[0,1] : \ f'(x) \text{ exists for some } x \in [0,1]\}.$$

*Then the set $S$ is of the first category in $C[0,1]$.*

The above theorem asserts that most continuous functions do not have derivatives at any point, or continuous functions that are differentiable at even one point in $[0,1]$ form a "sparse" subset. This is not intuitive, but an unexpected consequence of Baire's Theorem: most continuous functions are nowhere differentiable! For a construction of such a function we refer the reader to, e.g., [1], pp. 146–148.

## Inner Product Spaces

The concept of a vector space generalizes what we already know about addition and scalar multiplication on $\mathbb{R}$ or $\mathbb{R}^2$. However, if we want to talk about the concept of *length* or *angle* in a vector space, we need to study inner product spaces. To motivate the concept of inner product spaces, consider $x \in \mathbb{R}^2$ with initial point at the origin.

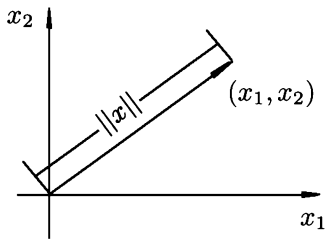



Figure 2.6: The norm of $x$

Then the vector $x = (x_1, x_2)$ has a *length* which is called the *norm* of $x$, denoted by $||x||$ (Figure 2.6) and defined as

$$||x|| = \sqrt{x_1^2 + x_2^2}.$$

An obvious generalization of the norm of $x \in \mathbb{R}^n$ is:

$$||x|| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

To utilize linearity in our discussion we introduce the dot product.

**Definition 56** (Dot Product)**.** The dot product of two vectors $x, y \in \mathbb{R}^n$ where $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_1, \ldots, y_n)$ is defined as

$$x \cdot y = x_1 y_1 + \cdots + x_n y_n = \sum_{i=1}^{n} x_i y_i.$$

This number relates to the norm by

$$||x||^2 = x \cdot x.$$

Note that $x \cdot x \geq 0$ for all $x \in \mathbb{R}^n$ with equality in the case $x = 0$. Furthermore, $x \cdot y = y \cdot x$ for all $x, y \in \mathbb{R}^n$ and if we keep $y \in \mathbb{R}^n$ fixed, we can define the linear map

$$\Phi : \mathbb{R}^n \to \mathbb{R} \quad \text{by} \quad \Phi(x) = x \cdot y.$$

An *inner product* is a generalization of a *dot product*. Thus the question we try to answer is if we consider the above properties of a dot product and define an inner product by abstracting these concepts, will we get all the properties of inner product spaces? It turns out that the answer is yes for **real** vector spaces. But if we want to define inner product for real as well as **complex** vector spaces we need to recall some concepts about complex numbers. Let $z = x + iy \in \mathbb{C}$ where $x, y \in \mathbb{R}$. The absolute value $|z|$ of $z$ and the complex conjugate $\overline{z}$ of a $z$ are defined as:

$$|z| = \sqrt{x^2 + y^2}, \quad \overline{z} = x - iy.$$

Note that

$$z\overline{z} = (x+iy)(x-iy) = x^2+y^2 = |z|^2.$$

Now, we are ready to define a complex norm.

**Definition 57** (Complex Norm)**.** For $z \in \mathbb{C}^n$ where $z = (z_1, z_2, \ldots, z_n)$, then the norm of $z$ is defined as

$$||z|| = \sqrt{|z_1|^2 + |z_2|^2 + \cdots + |z_n|^2},$$

where we need absolute values because $||z||$ is a non-negative number.

Note that

$$||z||^2 = |z_1|^2 + |z_2|^2 + \cdots + |z_n|^2 = z_1\overline{z_1} + z_2\overline{z_2} + \cdots + z_n\overline{z_n}.$$

So we can define an inner product of $z, w \in \mathbb{C}^n$.

**Definition 58** (Complex Inner Product)**.** For $z, w \in \mathbb{C}^n$, we define the complex inner product as

$$z \cdot w = z_1\overline{w_1} + z_2\overline{w_2} + \cdots + z_n\overline{w_n}.$$

**Example 65.** Let $z = (i, 1+i, 3)$ and $w = (1-i, 2, 4i)$. Let us find $||z||$, $z \cdot w$ and $d(z, w)$.

First,

$$\begin{aligned} ||z||^2 &= z_1\overline{z_1} + z_2\overline{z_2} + z_3\overline{z_3} = (i)(-i) + (1+i)(1-i) + 3 \cdot 3 \\ &= -i^2 + 1 - i + i - i^2 + 9 = 1 + 1 + 1 + 9 = 12, \end{aligned}$$

which implies $||z|| = \sqrt{12}$. Then

$$\begin{aligned} z \cdot w &= i(1+i) + (1+i)(2) + 3(-4i) = i + i^2 + 2 + 2i - 12i \\ &= 1 - 9i. \end{aligned}$$

We know that $d(z, w) = ||z - w||$, so we will first determine $z - w$.

$$z - w = (i, 1+i, 3) - (1-i, 2, 4i) = (-1+2i, -1+i, 3-4i).$$

So we have

$$\begin{aligned} ||z-w||^2 &= |-1+2i|^2 + |-1+i|^2 + |3-4i|^2 \\ &= [(-1)^2 + (2)^2] + [(-1)^2 + (1)^2] + [(3)^2 + (-4)^2] \\ &= 5 + 2 + 25 = 32, \end{aligned}$$

which implies $||z-w|| = \sqrt{32} = 4\sqrt{2}$.

**Definition 59** (Inner Product)**.** An inner product on a real or complex vector space $V$ is a function

$$\langle\ ,\ \rangle : V \times V \to \mathbb{F}$$

(the field $\mathbb{F}$ could be $\mathbb{R}$ or $\mathbb{C}$) and this function has the following properties:

a) $\langle v, v\rangle \geq 0$ for all $v \in V$ (positivity).

b) $\langle v, v\rangle = 0 \Leftrightarrow v = 0$ (definiteness).

c) $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$ for all $u, v, w \in V$ (additivity in first slot).

d) $\langle \alpha v, w\rangle = \alpha\langle v, w\rangle$ for all $\alpha \in \mathbb{F}$ and all $v, w \in V$ (homogeneity in first slot).

e) $\langle v, w\rangle = \overline{\langle w, v\rangle}$ for all $v, w \in V$ (conjugate symmetry).

Note that if we are in a *real* inner product space, then $\overline{\langle w, v\rangle} = \langle w, v\rangle$ because every real number is equal to its complex conjugate. Then for condition $e$), we simply write

$$\langle v, w\rangle = \langle w, v\rangle,$$

for all $v, w \in V$.

**Example 66.** The following are examples of inner products on various spaces. It is routine to check these functions actually define an inner product.

a) For $x, y \in \mathbb{R}^2$,

$$\langle x, y\rangle = 2x_1y_1 + 3x_2y_2$$

defines an inner product.

b) For $x, y \in \mathbb{R}^n$,

$$\langle x, y\rangle = x_1y_1 + x_2y_2 + \cdots + x_ny_n$$

defines an inner product on $\mathbb{R}^n$.

**Example 67.** Let $C[a, b]$ be the space of all continuous functions on $[a, b]$, namely,

$$C[a, b] := \{f :\ f : [a, b] \to \mathbb{R},\ f \text{ is continuous}\}.$$

Consider $f, g \in C[a, b]$. We can define an inner product on $C[a, b]$ as:

$$\langle f, g\rangle = \int_a^b f(x)g(x)dx.$$

Check the conditions given in Definition 59.

**Example 68.** Let $P_2(\mathbb{R})$ be the space of all polynomials of real coefficients with degree at most 2. Given $p(x), g(x) \in P_2(\mathbb{R})$ where $p(x) = a_0 + a_1x + a_2x^2$ and $g(x) = b_0 + b_1x + b_2x^2$ then

$$\langle p, g\rangle = a_0b_0 + a_1b_1 + a_2b_2,$$

defines an inner product on $P_2(\mathbb{R})$.

## Orthogonal Decomposition

**Definition 60** (Orthogonal)**.** Two vectors $u, v \in V$ are called *orthogonal* if

$$\langle u, v \rangle = 0.$$

Note that in the above definition order does not matter, since $\langle u, v \rangle = 0$ if and only if $\langle v, u \rangle = 0$. Furthermore, the zero vector, 0, is orthogonal to every vector and 0 is the only vector that is orthogonal to itself. Some mathematicians use the term *perpendicular* instead of *orthogonal*.

**Theorem 67** (Pythagorean Theorem)**.** *If $u$ and $v$ are orthogonal vectors in $V$, then*

$$||u + v||^2 = ||u||^2 + ||v||^2.$$

*Proof.* Suppose $u$ and $v$ are orthogonal vectors in $V$; then using the definition of the norm and the properties of the inner product we have:

$$\begin{aligned} ||u+v||^2 &= \langle u+v, u+v \rangle = \langle u, u \rangle + \langle u, v \rangle + \langle v, u \rangle + \langle v, v \rangle \\ &= ||u||^2 + \langle u, v \rangle + \overline{\langle u, v \rangle} + ||v||^2 \\ &= ||u||^2 + 2\mathrm{Re}\langle u, v \rangle + ||v||^2 \\ &= ||u||^2 + ||v||^2. \quad \square \end{aligned}$$

**Remark 37.** The converse of the Pythagorean theorem holds in real inner product spaces. Note that in the above proof of Pythagorean theorem, the equality in Theorem 67 holds if and only if

$$\langle u, v \rangle + \langle v, u \rangle = 2\mathrm{Re}\langle u, v \rangle = 0.$$

However, if we are in a **real** inner product space, then

$$\langle u, v \rangle = \langle v, u \rangle = 0.$$

Thus the converse of the Pythagorean theorem holds in real inner product spaces.

We are now ready to focus on orthogonal decomposition. Let $u, v \in V$. Suppose we want to write $u = \alpha v + w$ where $\alpha$ is a scalar and $w$ is a vector orthogonal to $v$ (Figure 2.7). Expressing this in another way, we can write

$$u = \alpha v + (u - \alpha v).$$

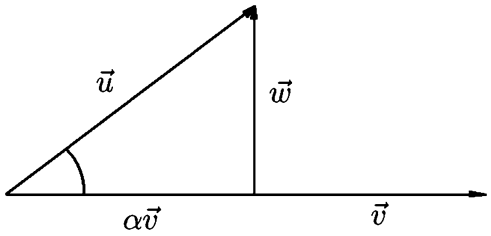



Figure 2.7: Orthogonal decomposition

To find this scalar $\alpha$, notice that we need to choose $\alpha$ so that $v \perp (u - \alpha v)$ which means that

$$\begin{aligned} \langle v, u - \alpha v \rangle &= 0 \\ \Rightarrow \langle v, u \rangle - \alpha ||v||^2 &= 0 \\ \Rightarrow \alpha &= \frac{\langle v, u \rangle}{||v||^2}. \end{aligned}$$

By assuming that $v \neq 0$ we can write

$$u = \frac{\langle v, u \rangle}{||v||^2} v + \left( u - \frac{\langle v, u \rangle}{||v||^2} v \right).$$

Thus, we were able to write $u$ as a scalar multiple of $v$ plus a vector $w = u - \frac{\langle v, u \rangle}{||v||^2} v$ which is perpendicular to $v$. The vector $\alpha v$ is often named the *projection* of $u$ onto $v$ and defined as:

$$\text{Proj}_v(u) = \frac{\langle v, u \rangle}{||v||^2} v.$$

We will use this decomposition to prove the Cauchy–Schwarz–Bunyakovsky (CSB) inequality.

## The Cauchy–Schwarz–Bunyakovsky Inequality

The Cauchy–Schwarz–Bunyakovsky inequality is one of the most useful inequalities one can encounter in different settings. It has been used in linear algebra, analysis, and probability theory as well as other disciplines such as physics. In physics, the general formulation of the Heisenberg Uncertainty Principle, a fundamental part of quantum mechanics, is derived by using the CSB inequality in Hilbert spaces. The CSB inequality also has many generalizations: Hölder's Inequality for $L^p$-spaces is one such example. The inequality for sums was published by Augustin-Louis Cauchy in 1821. The integral analog of Cauchy's inequality is known as the Schwarz inequality and it was presented by Herman Amandus Schwarz in 1885. It appears that the same result had been independently formulated in 1859 by a Russian mathematician, Victor Yacoulevich Bunyakovsky, and thus also bears his name. We refer the reader to J. M. Steele's text [48] *The Cauchy–Schwartz Master Class: An Introduction to the Art of Mathematical Inequalities* for further information. Now, we proceed with the result.

**Theorem 68** (Cauchy–Schwarz–Bunyakovsky Inequality)**.** *Let $u$ and $v$ be vectors in an inner product space $V$. Then*

$$|\langle u, v \rangle| \leq ||u|| \cdot ||v||,$$

*with equality holding if and only if one of $u$, $v$ is a scalar multiple of the other.*

*Proof.* If $v = 0$, then the inequality is actually an equality since

$$|\langle u, 0\rangle| = 0 = ||u|| \cdot ||0||.$$

So, let $v \neq 0$ and consider the orthogonal decomposition

$$u = \frac{\langle v, u\rangle}{||v||^2} v + w.$$

By applying the Pythagorean Theorem,

$$\begin{aligned} ||u||^2 &= \left|\left| \frac{\langle v, u\rangle}{||v||^2} v \right|\right|^2 + ||w||^2 \\ &= \frac{|\langle v, u\rangle|^2}{||v||^2} + ||w||^2 \\ &\geq \frac{|\langle v, u\rangle|^2}{||v||^2}. \end{aligned}$$

By multiplying both sides by $||v||^2$ and taking the square roots of both sides, we are left with the CSB inequality. □

## Some Applications of CSB Inequality

The first application of the CSB inequality is the *triangle inequality*, which is also extremely useful.

**Theorem 69** (The Triangle Inequality)**.** *Let $u$ and $v$ be vectors in an inner product space $V$. Then*

$$||u + v|| \leq ||u|| + ||v||.$$

*Proof.* Let $u, v \in V$.

$$\begin{aligned} ||u + v||^2 &= \langle u + v, u + v\rangle \\ &= \langle u, u\rangle + \langle v, v\rangle + \langle u, v\rangle + \langle v, u\rangle \\ &= ||u||^2 + ||v||^2 + \langle u, v\rangle + \overline{\langle u, v\rangle} \\ &= ||u||^2 + ||v||^2 + 2\mathrm{Re}\langle u, v\rangle \\ &\leq ||u||^2 + ||v||^2 + 2|\langle u, v\rangle| \\ &\leq ||u||^2 + ||v||^2 + 2||u|| \cdot ||v|| = (||u|| + ||v||)^2, \end{aligned}$$

where we have used the CSB inequality

$$2|\langle u, v\rangle| \leq 2||u|| \cdot ||v||.$$

This implies that

$$||u + v|| \leq ||u|| + ||v||.$$ □

If we consider $x, y \in \mathbb{R}^n$ with the Euclidean norm, the CSB inequality takes on the form:

$$\sum_{i=1}^{n} |x_i y_i| \leq \left( \sum_{i=1}^{n} |x_i|^2 \right)^{1/2} \left( \sum_{i=1}^{n} |y_i|^2 \right)^{1/2},$$

or in our notation we have:

**Theorem 70** (Cauchy's Inequality)**.** *Let $x, y \in \mathbb{R}^n$ with the Euclidean norm. Then*

$$\sum_{i=1}^{n} |x_i y_i| \leq ||x|| \cdot ||y||.$$

Note that we may suppose that $x \neq 0$, $y \neq 0$ in the above Theorem 70; there is nothing to show if either is 0.

**Remark 38** (Schwarz Inequality)**.** The integral analog of Cauchy's inequality is known as the Schwarz inequality. To explain Schwarz Inequality we need to define the set $L^2[a, b]$ and understand the inner product on this space. $L^2[a, b]$ denote the space of square Lebesgue integrable functions, Lebesgue integrable functions will be defined later in the text, see Definition 98.

At this point we just want to observe that for $f, g \in L^2[a, b]$ if the inner product given as below, then we have

$$\begin{aligned} |\langle f, g \rangle| &= \left| \int_a^b f(x) g(x) dx \right| \leq \int_a^b |f(x)| \cdot |g(x)| dx \\ &\leq \left( \int_a^b |f(x)|^2 \right)^{1/2} \left( \int_a^b |g(x)|^2 \right)^{1/2}. \end{aligned}$$

Using the CSB inequality, we can prove interesting inequalities.

Let $x$ and $y$ be non-negative real numbers. Applying the CSB inequality to

$$u = \begin{bmatrix} \sqrt{x} \\ \sqrt{y} \end{bmatrix} \quad v = \begin{bmatrix} \sqrt{y} \\ \sqrt{x} \end{bmatrix}$$

yields a useful inequality:

**Theorem 71** (AM-GM Inequality)**.** *Let $x, y \in \mathbb{R}_+$ (positive real numbers). Then,*

$$\sqrt{xy} \leq \frac{x + y}{2}.$$

*Proof.* We know that

$$\begin{aligned} |\langle u, v \rangle| &= 2\sqrt{xy}, \\ ||u||^2 &= x + y, \\ ||v||^2 &= x + y, \end{aligned}$$

and $x$ and $y$ are non-negative implying that $||u|| = ||v|| = \sqrt{x+y}$. Therefore,

$$2\sqrt{xy} = ||\langle u, v\rangle|| \leq ||u|| \cdot ||v|| = \sqrt{x+y}\sqrt{x+y} = x + y.$$

The right side of AM-GM inequality is the familiar arithmetic mean of $x, y$ and the left side is the geometric mean. □

**Corollary 13.** *The AM-GM inequality holds for $n$ non-negative variables, $x_1$, $x_2, \ldots, x_n$. It has the form:*

$$\sqrt[n]{x_1 x_2 \cdots x_n} \leq \frac{x_1 + x_2 + \cdots + x_n}{n},$$

*with equality if and only if $x_1 = x_2 = \cdots = x_n$.*

**Lemma 8** (Minkowski's Inequality)**.** *Let $x, y \in \mathbb{R}^n$ or $\mathbb{C}^n$. Then*

$$||x + y|| \leq ||x|| + ||y||,$$

*or equivalently*

$$\left(\sum_{i=1}^{n} |x_i + y_i|^2\right)^{1/2} \leq \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2} + \left(\sum_{i=1}^{n} |y_i|^2\right)^{1/2}.$$

*Proof.* Using Cauchy's inequality,

$$\begin{aligned} ||x+y||^2 &= \sum_{i=1}^{n} |x_i + y_i| \cdot |x_i + y_i| \leq \sum_{i=1}^{n} |x_i + y_i|(|x_i| + |y_i|) \\ &= \sum_{i=1}^{n} |x_i + y_i| \cdot |x_i| + \sum_{i=1}^{n} |x_i + y_i| \cdot |y_i| \\ &\leq ||x+y|| \cdot ||x|| + ||x+y|| \cdot ||y|| = ||x+y||(||x|| + ||y||). \end{aligned}$$

This implies that

$$||x+y||^2 \leq ||x+y||(||x|| + ||y||).$$

If $||x + y|| = 0$, then the above lemma is trivially true. Otherwise, Minkowski's inequality follows from the inequality last written by dividing through by $||x + y||$. □

**Remark 39.** Observe that Lemma 8 is a special case of the triangle inequality given in Theorem 69. It illustrates the triangle inequality for $x, y \in \mathbb{R}^n$ or $\mathbb{C}^n$.

Suppose a given inner product space has been normed by taking

$$||w|| = \sqrt{\langle w, w\rangle}.$$

We say a sequence $\{x_n\}$ converges to $x$ when $||x_n - x|| \to 0$. Since every inner product gives rise to a norm, we can talk about convergent sequences or Cauchy sequences in an inner product space as well. The following is such an example:

**Theorem 72.** *If $\{x_n\}$ and $\{y_n\}$ are sequences in an inner product space, which converge to $x$ and $y$, respectively, then $\{\langle x_n, y_n\rangle\}$ is convergent sequence with limit $\langle x, y\rangle$.*

*Proof.* To prove $\langle x_n, y_n\rangle \to \langle x, y\rangle$, we write

$$\begin{aligned}\langle x_n, y_n\rangle - \langle x, y\rangle &= \langle x_n, y_n\rangle - \langle x_n, y\rangle + \langle x_n, y\rangle - \langle x, y\rangle \\ &= \langle x_n, y_n - y\rangle + \langle x_n - x, y\rangle,\end{aligned}$$

so that by using the CSB inequality we obtain:

$$\begin{aligned}|\langle x_n, y_n\rangle - \langle x, y\rangle| &\leq |\langle x_n, y_n - y\rangle| - |\langle x_n - x, y\rangle| \\ &\leq ||x_n||||y_n - y|| + ||x_n - x||||y||.\end{aligned}$$

Every convergent sequence is bounded, so $||x_n|| \leq K$ for some constant $K$ and for all $n$. Moreover $||x_n - x|| \to 0$ and $||y_n - y|| \to 0$, hence, as claimed, $\langle x_n, y_n\rangle \to \langle x, y\rangle$. □

## Orthonormal Vectors

A list $e_1, \ldots, e_n$ of vectors in $V$ is called *orthonormal* if

$$\langle e_j, e_k\rangle = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{if } j \neq k \end{cases}.$$

**Example 69.** The standard basis $\{e_1, e_2, \ldots, e_n\}$ in $\mathbb{R}^n$ is an orthonormal list, where by $e_i$ we mean a vector in $\mathbb{R}^n$ with $i$-th component equal to 1 and all other components are zero. For example $\{e_1 = (1, 0, 0), e_2 = (0, 1, 0), e_3 = (0, 0, 1)\}$ is an orthonormal list in $\mathbb{R}^3$. Another example of an orthonormal list in $\mathbb{R}^3$ is

$$\left\{\left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right), \left(\frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right), \left(\frac{-1}{\sqrt{6}}\ \frac{-1}{\sqrt{6}}, \frac{2}{\sqrt{6}}\right)\right\}.$$

Orthonormal lists are easy to work with because an orthonormal list of the right length is an orthonormal basis of $V$. Observe that an orthonormal list is linearly independent because if we set $a_1 e_1 + \cdots + a_m e_m = 0$, then the fact that

$$||a_1 e_1 + \cdots + a_m e_m||^2 = |a_1|^2 + \cdots + |a_m|^2$$

for all scalars $a_1, \ldots, a_m$ yields that all $a_j$'s are zero. Thus $\{e_1, e_2, \ldots, e_m\}$ is linearly independent. Another important property of orthonormal bases is the fact that one can write a given vector $v \in V$ as a linear combination of orthonormal basis elements, namely:

**Theorem 73.** *Let $v \in V$ and $e_1, \ldots, e_n$ be an orthonormal basis for $V$. Then*

$$v = \langle v, e_1\rangle e_1 + \cdots + \langle v, e_n\rangle e_n$$

*and*

$$||v||^2 = |\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_n\rangle|^2.$$

*Proof.* Because $e_1, \ldots, e_n$ is an orthonormal basis for $V$, there exist scalars $a_1, \ldots, a_n$ such that

$$v = a_1 e_1 + \cdots + a_n e_n.$$

Taking the inner product of both sides of this equation with $e_j$ yields $\langle v, e_j \rangle = a_j$ which proves the first equation in the above theorem. To show

$$||v||^2 = |\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_n \rangle|^2$$

set $v = a_1 e_1 + \cdots + a_n e_n$ and use the fact that $||a_1 e_1 + \cdots + a_m e_m||^2 = |a_1|^2 + \cdots + |a_m|^2$. □

Now that we have some idea about how useful it is to have orthonormal bases, how do we go about finding them? The answer to this question is given by the Gram–Schmidt Procedure.

**Theorem 74** (Gram–Schmidt Procedure)**.** *Let $\{v_1, v_2, \ldots, v_m\}$ be a linearly independent set of vectors in an inner product space. Set*

- $e_1 = \dfrac{v_1}{||v_1||}$.
- *For $j = 2, 3, \ldots, m$ define:*

$$e_j = \frac{v_j - \langle v_j, e_1 \rangle e_1 - \cdots - \langle v_j, e_{j-1} \rangle e_{j-1}}{||v_j - \langle v_j, e_1 \rangle e_1 - \cdots - \langle v_j, e_{j-1} \rangle e_{j-1}||}.$$

*Then $e_1, e_2, \ldots, e_m$ is an orthonormal set such that for $j = 1, \ldots, m$ we have*

$$\mathit{span}(e_1, \ldots, e_j) = \mathit{span}(v_1, \ldots, v_j).$$

The proof of the above theorem can be found in any standard textbook on linear algebra. In the following we give a couple of examples.

**Example 70.** Let $\{v_1, v_2, v_3\}$ be a list of linearly independent vectors in $\mathbb{R}^4$, where $v_1 = (1, 0, 1, 0)$, $v_2 = (1, 1, 1, 1)$ and $v_3 = (0, 1, 2, 1)$. To find the orthonormal list we set

$$e_1 = \frac{v_1}{||v_1||} = \frac{(1, 0, 1, 0)}{\sqrt{2}},$$

$$e_2 = \frac{v_2 - \langle v_2, e_1 \rangle e_1}{||v_2 - \langle v_2, e_1 \rangle e_1||} = \frac{(0, 1, 0, 1)}{\sqrt{2}},$$

and finally

$$e_3 = \frac{v_3 - \langle v_3, e_1 \rangle e_1 - \langle v_3, e_2 \rangle e_2}{||v_3 - \langle v_3, e_1 \rangle e_1 - \langle v_3, e_2 \rangle e_2||} = \frac{(-1, 0, 1, 0)}{\sqrt{2}}.$$

Observe that each $e_j$ for $j = 1, 2, 3$ has length 1 and

$$\langle e_1, e_2 \rangle = \langle e_1, e_3 \rangle = \langle e_2, e_3 \rangle = 0,$$

i.e., $\{e_1, e_2, e_3\}$ is an orthonormal list.

**Example 71.** Consider the finite dimensional vector space $\mathcal{P}_2(\mathbb{R})$, the set of all polynomials with coefficients in $\mathbb{R}$ and degree at most 2. Suppose the inner product on $\mathcal{P}_2(\mathbb{R})$ is given by:

$$\langle p, q\rangle = \int_{-1}^{1} p(x)\, q(x)\, dx.$$

We know that the set $\{1, x, x^2\}$ is a linearly independent list; in fact it is a basis for $\mathcal{P}_2(\mathbb{R})$. To find an orthonormal basis for this space we apply the Gram–Schmidt process and because $||1||^2 = \int_{-1}^{1} 1 \cdot 1\, dx = 2$ we have

$$e_1 = \frac{v_1}{||v_1||} = \frac{1}{||1||} = \frac{1}{\sqrt{2}}.$$

To find $e_2$, we find $\langle v_2, e_1\rangle = \int_{-1}^{1} x\frac{1}{\sqrt{2}} = \int_{-1}^{1} x\frac{1}{\sqrt{2}}\, dx = 0$ and thus

$$e_2 = \frac{v_2 - \langle v_2, e_1\rangle e_1}{||v_2 - \langle v_2, e_1\rangle e_1||} = \frac{x}{\sqrt{\frac{2}{3}}} = \sqrt{\frac{3}{2}}x.$$

To find $e_3$ we first set $\langle v_3, e_1\rangle = \int_{-1}^{1} x^2\sqrt{\frac{1}{2}}\, dx$ and $\langle v_3, e_2\rangle = \int_{-1}^{1} x^2\sqrt{\frac{3}{2}}x\, dx$ and

$$v_2 - \langle v_2, e_1\rangle e_1 - \langle v_2, e_2\rangle e_2 = x^2 - \langle x^2, e_1\rangle e_1 - \langle x^2, e_2\rangle e_2 = x^2 - \frac{1}{3}.$$

Therefore $e_3 = \dfrac{x^2 - \frac{1}{3}}{||x^2 - \frac{1}{3}||} = \sqrt{\dfrac{45}{8}}\left(x^2 - \dfrac{1}{3}\right)$. Note that the orthonormal set consists of $\{e_1, e_2, e_3\}$ where

$$\begin{aligned} e_1 &= \sqrt{\frac{1}{2}} \\ e_2 &= \sqrt{\frac{3}{2}}\, x \\ e_3 &= \sqrt{\frac{45}{8}}\left(x^2 - \frac{1}{3}\right). \end{aligned}$$

**Remark 40.** Many orthonormal sets of great interest and importance in analysis can be obtained by applying the Gram–Schmidt processes. One such orthonormal set is in $L^2[-1, 1]$, where by $L^2[-1, 1]$ we mean the space of square Lebesgue integrable functions (see Definition 98 for Lebesgue integrability later in the text).

If we want to obtain an orthonormal set in $L^2[-1, 1]$, we start with polynomials

$$x_0(t) = 1, x_1(t) = t, \ldots, x_j(t) = t^j, \ldots \quad t \in [-1, 1].$$

This sequence of functions is linearly independent (why?) and thus we can apply the Gram–Schmidt process to obtain an orthonormal set. It can be shown that

$$e_n(t) = \sqrt{2n + 1/2}\, P_n(t) \qquad n = 0, 1, \ldots$$

where $P_n$ is called the *Legendre polynomial* of order $n$ given by the formula

$$P_n(t) = \frac{1}{2^n n!} \frac{d^n}{dt^n} \left[ (t^2 - 1)^n \right].$$

By applying the binomial theorem to $(t^2 - 1)^n$ and differentiating the result $n$ times term by term, one can obtain an equivalent form

$$P_n(t) = \sum_{j=0}^{N} \frac{(2n - 2j)!}{2^n j!(n - j)!(n - 2j)!}\, t^{n-2j},$$

where if $n$ is even we take $N = n/2$ and in case $n$ is odd $N = (n - 1)/2$. If we list the first few Legendre polynomials, we see that

$$\begin{aligned} P_0(t) &= 1, \\ P_1(t) &= t, \\ P_2(t) &= \frac{1}{2}(3t^2 - 1), \\ P_3(t) &= \frac{1}{2}(5t^3 - 3t), \\ P_4(t) &= \frac{1}{8}(35t^4 - 30t^2 + 3). \end{aligned}$$

Note that $P_n(t)$ is a polynomial of degree $n$ and it is an orthonormal set in $L^2[-1, 1]$. Legendre polynomials are used in analysis and approximation theory; they also have applications in physics, where they are used in the determination of wave functions.

Another important application of the Gram–Schmidt process is the following well known result, known as the Riesz representation theorem.

## The Riesz Representation Theorem

The Riesz representation theorem gives us a characterization of the set of all linear functionals on a finite dimensional inner product space. The advantage of this theorem is that it gives an expression for a whole class of linear functionals. These ideas motivated many other results in functional analysis and in advanced analysis. We start by defining linear functionals.

**Definition 61.** A *linear functional* $f$ on a vector space $X$ is a linear map $X \to \mathbb{F}$ where $\mathbb{F}$ is the scalar field of $X$. If $X$ is a real inner product space then $\mathbb{F} = \mathbb{R}$, and $\mathbb{F} = \mathbb{C}$ if $X$ is complex.

**Definition 62.** A linear functional on $X$ is said to be *bounded* if there exists a constant $M \geq 0$ such that

$$|f(x)| \leq M||x||$$

for all $x \in X$.

For a given bounded linear functional $f$, we may find many $M$'s satisfying the above condition. If it is valid for one $M$, it is valid for $M_1 > M$ also. The next theorem shows that the condition of boundedness is equivalent to continuity. For the proof of the following theorem we refer to, e.g., [35], p. 97.

**Theorem 75.** *A linear functional $f$ on $X$ is continuous if and only if $f$ is bounded.*

In the following we list some examples of linear functionals:

a) $f : \mathbb{R}^n \to \mathbb{R}$, where $f(x) = \sum_{k=1}^{n} a_k x_k$, $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ and $(a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$ is fixed.

b) $f : C[a,b] \to \mathbb{R}$, where $f(x) = \int_a^b x(t)\, dt$, $x \in C[a,b]$.

Note that for a normed space $X$ and $x \in X$, $f : X \to \mathbb{R}$ defined as $f(x) = ||x||$ is a functional on $X$ which is not linear.
Suppose $X$ is a finite dimensional, nonzero inner product space over $\mathbb{F}$. Let $v \in X$ be a fixed vector, then for $x \in X$ the mapping $f$ given by:

$$f(x) = \langle x, v \rangle$$

is a linear functional. Linearity follows from

$$\begin{aligned} f(a_1x_1 + a_2x_2) &= \langle a_1x_1 + a_2x_2, v \rangle \\ &= a_1\langle x_1, v \rangle + a_2\langle x_2, v \rangle \\ &= a_1 f(x_1) + a_2 f(x_2). \end{aligned}$$

The Riesz theorem states that there are in fact no other types of linear functionals on a finite dimensional inner product space, thus any linear functional on the space $X$ must have the form $f(x) = \langle x, w \rangle$ for some unique point $w$ in $X$.

**Theorem 76** (The Riesz representation theorem)**.** *Let $X$ be a finite dimensional inner product space and let $f$ be a linear functional on $X$. Then there exists a unique point $v \in X$ such that $f(x) = \langle x, v \rangle$ for all $x \in X$.*

*Proof.* First we show there exists a vector $v \in X$ such that $f(x) = \langle x, v \rangle$ for every $x \in X$. Let $e_1, \ldots, e_n$ be an orthonormal basis for $X$. Then,

$$\begin{aligned} f(x) &= f(\langle x, e_1 \rangle e_1 + \cdots + \langle x, e_n \rangle e_n) \\ &= \langle x, e_1 \rangle f(e_1) + \cdots + \langle x, e_n \rangle f(e_n) \\ &= \langle x, \overline{f(e_1)} e_1 + \cdots + \overline{f(e_n)} e_n \rangle. \end{aligned}$$

Setting $v = \overline{f(e_1)}e_1 + \cdots + \overline{f(e_n)}e_n$ we have $f(x) = \langle x, v\rangle$ for every $x \in X$ as desired. Now we prove that this vector is unique. Suppose we have two vectors $v_1, v_2 \in X$ such that

$$f(x) = \langle x, v_1\rangle = \langle x, v_2\rangle$$

for every $x \in X$. Then

$$0 = \langle x, v_1\rangle - \langle x, v_2\rangle = \langle x, v_1 - v_2\rangle$$

for every $x \in X$. By taking $x = v_1 - v_2$ we show that $v_1 - v_2 = 0$, in other words $v_1 = v_2$. □

Suppose $X$ is finite dimensional and $f$ is a linear functional on $X$, then by the above proof the vector $v$ satisfies $f(x) = \langle x, v\rangle$ where

$$v = \overline{f(e_1)}e_1 + \cdots + \overline{f(e_n)}e_n.$$

The right side of this equation seems to depend on the functional $f$ as well as the orthonormal basis $e_1, e_2, \ldots, e_n$. However the Riesz representation theorem tells us that $v$ is uniquely determined by $f$. Hence the right side of the equation is the same regardless of which orthonormal basis is taken for $X$. The following is a simple consequence of this theorem.

**Theorem 77.** *All linear functionals on finite dimensional inner product spaces are bounded.*

*Proof.* If $f$ is a linear functional on $X$, then there is a point $v \in X$ such that $f$ has the form $f(x) = \langle x, v\rangle$ for all $x \in X$. But then by the CSB inequality

$$|f(x)| = |\langle x, v\rangle| \leq ||x||||v||,$$

so $f$ is bounded. □

Note that we can say even more; if we set $||f|| = \sup_{||x||=1} |f(x)|$, then $||f|| \leq ||v||$. In fact $||f|| = ||v||$. This is seen by noting that

$$|f(v)| = |\langle v, v\rangle| = ||v||^2$$

so that we cannot have $|f(x)| < ||x||||v||$ for all $x \in X$.

**Remark 41.** There are many other versions of the Riesz representation theorem giving similar results in other spaces. Even though we proved this theorem for finite dimensional inner product spaces, for a complete inner product space called a *Hilbert space*, the Riesz representation theorem gives rise to the quite important notion of an operator *adjoint* to a given operator.

**Theorem 78.** *If $f$ is a bounded linear functional on a Hilbert space $H$, then there is a unique point $v \in H$ such that $f(x) = \langle x, v\rangle$ for all $x \in H$.*

The proof of this theorem is considerably more involved (for proof consult [35], pp. 188–190), but note that here we require the functional to be bounded, whereas in the former case boundedness is obtained as a consequence.

## Normed Vector Spaces

**Definition 63** (Norm on a Vector Space)**.** A norm on a real or complex vector space $V$ is a function

$$||\cdot|| : V \to [0, \infty),$$

satisfying:

a) $0 \leq ||x|| < \infty$ for all $x \in V$.

b) $||x|| = 0 \Leftrightarrow x = 0$ (the zero vector in $V$).

c) $||\alpha x|| = |\alpha| \cdot ||x||$ for any scalar $\alpha$ and any $x \in V$.

d) $||x + y|| \leq ||x|| + ||y||$ for all $x, y \in V$ (triangle inequality).

**Definition 64** (Normed vector space)**.** The pair $(V, ||\cdot||)$ consisting of a real or complex vector space $V$ together with a norm on $V$ is called a *normed vector space.*

Any norm induces a metric on $V$ by setting:

$$d(x, y) = ||x - y||.$$

**Remark 42.** Not all metrics on a vector space come from a norm because if a metric is induced by a norm, then it must have the following properties:

$$\begin{aligned} d(x + a, y + a) &= ||x+a-(y+a)||=||x-y||=d(x, y) \text{ (translation invariant)} \\ d(\alpha x, \alpha y) &= ||\alpha x-\alpha y||=|\alpha|d(x, y), \end{aligned}$$

for all $x, y, a \in V$. So, there are metrics which are not derived from a norm. For example, if one consider the vector space $V$ as the space of all sequences $\{x_i\}$ with the metric

$$d(x, y) = \sum_{j=1}^{\infty} \left(\frac{1}{2^j}\right) \frac{|x_j - y_j|}{1 + |x_j - y_j|}.$$

Then, this metric (which is called the Fréchet metric) cannot be obtained from a norm (see that translation invariance fails).

The following are examples of norms on various spaces. Using Definition 63, one can check easily that these functions actually define a norm.

**Example 72.**

$$||x|| = |x|$$

is a norm on $\mathbb{R}$.

**Example 73.** Each of the following defines a norm on $\mathbb{R}^n$:

$$\begin{aligned} ||x||_1 &= \sum_{i=1}^{n} |x_i|, \\ ||x||_2 &= \left( \sum_{i=1}^{n} |x_i|^2 \right)^{1/2}, \\ ||x||_\infty &= \max_{1 \le i \le n} |x_i|. \end{aligned}$$

In general, for any $1 \le p < \infty$, the expression

$$||x||_p = \left( \sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

defines a norm on $\mathbb{R}^n$.

Note that $||x||_2$ is the usual Euclidean norm.

**Example 74.** Each of the following defines a norm on $C[a, b]$

$$\begin{aligned} ||f||_1 &= \int_a^b |f(t)| dt, \\ ||f||_2 &= \left( \int_a^b |f(t)|^2 dt \right)^{1/2}, \\ ||f||_\infty &= \max_{a \le t \le b} |f(t)|. \end{aligned}$$

**Example 75.** If $(X, ||\cdot||)$ is a normed vector space and $Y$ is a linear subspace of $X$, then $Y$ is also normed by $||\cdot||$. In other words, the restriction of $||\cdot||$ to $Y$ defines a norm on $Y$.

**Example 76** (Function Space)**.** Let $C[a, b]$ represent the collection of continuous real-valued functions $f : [a, b] \to \mathbb{R}$, then

$$||f|| := \sup_{x \in [a,b]} |f(x)|$$

defines a norm on $C[a, b]$. Note that $f$ being continuous on the compact set $[a, b]$ implies it is a bounded function by the Extreme Value Theorem. Furthermore, $||f|| = 0$ if and only if $f(x) = 0$ for each $x \in [a, b]$. To show the triangle inequality holds for the above norm observe:

$$||f + g|| = \sup_{x \in [a,b]} |f(x) + g(x)| \le \sup_{x \in [a,b]} |f(x)| + \sup_{x \in [a,b]} |g(x)| = ||f|| + ||g||.$$

**Definition 65** (Sequence Space)**.** For $1 \leq p < \infty$ we define the sequence space

$$\ell^p = \left\{ x = \{x_i\} : \sum_{i=1}^{\infty} |x_i|^p < \infty \right\}.$$

For $1 \leq p < \infty$ we define the norm

$$||x||_p = \left( \sum_{i=1}^{\infty} |x_i|^p \right)^{1/p}.$$

For the space $\ell^\infty$ we define the norm

$$||x||_\infty = \sup |x_n|.$$

Note that the space $\ell^2$ is of particular importance (why?).

## Parallelogram Equality and the Norm

A norm in an inner product space satisfies the important parallelogram equality (Figure 2.8).

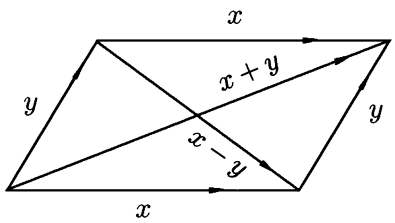



Figure 2.8: Parallelogram equality

**Theorem 79** (Parallelogram Equality)**.** *For $x, y$ in an inner product space*

$$||x+y||^2 + ||x-y||^2 = 2(||x||^2 + ||y||^2).$$

*Proof.* If $x, y \in X$, then

$$\begin{aligned} ||x+y||^2 + ||x-y||^2 &= \langle x+y, x+y\rangle + \langle x-y, x-y\rangle \\ &= ||x||^2 + ||y||^2 + \langle x, y\rangle + \langle y, x\rangle \\ &\qquad +||x||^2 + ||y||^2 - \langle x, y\rangle - \langle y, x\rangle \\ &= 2(||x||^2 + ||y||^2) \end{aligned}$$

as desired. □

If a norm does not satisfy the parallelogram equality, it cannot be obtained from an inner product. Thus not all normed spaces are inner product spaces.

**Example 77.** The space $\ell^p$ with $p \neq 2$ is a normed space but not an inner product space.

*Proof.* Recall

$$\ell^p = \left\{ x = \{x_i\} : \sum_{i=1}^{\infty} |x_i|^p < \infty \right\}.$$

Consider $x = (1, 1, 0, 0, \dots) \in \ell^p$ and $y = (1, -1, 0, 0, \dots) \in \ell^p$. Then

$$\begin{aligned} x + y &= (2, 0, 0, 0, \dots) \\ x - y &= (0, 2, 0, 0, \dots) \\ ||x + y||_p^p &= \sum_{i=1}^{\infty} |x_i|^p = |2|^p \\ \implies ||x + y||_p &= (2^p)^{1/p} = 2. \end{aligned}$$

Similarly, $||x - y||_p = 2$. On the other hand,

$$||x||_p^p = \sum_{i=1}^{\infty} |x_i|^p = 1^p + 1^p = 2 \implies ||x||_p = 2^{1/p} = ||y||_p.$$

So the parallelogram equality for this norm does not hold

$$\begin{aligned} ||x + y||_p^2 + ||x - y||_p^2 &= 2(||x||_p^2 + ||y||_p^2) \\ 4 + 4 &\neq 2\left(2^{1/p} + 2^{1/p}\right) \end{aligned}$$

unless $p = 2$. □

## Matrix Norms

We can form norms for matrices the same way that we defined norms for vectors in $\mathbb{R}^n$. After all, the vector space $M_{mn}$ of all $m \times n$ matrices is isomorphic to $\mathbb{R}^{mn}$. Thus the properties a) to c) of a norm will also hold in the setting of matrices. It turns out that, for matrices, the norms that are most useful satisfy an additional property d).

**Definition 66.** A *matrix norm* on $M_{nn}$ is a mapping that associates with each $n \times n$ matrix $A$ a real number $||A||$, called the norm of $A$, such that the following hold for all $n \times n$ matrices $A$ and $B$ and all scalars $c$.

a) $||A|| \geq 0$ and $||A|| = 0$ if and only if $A = 0$.

b) $||cA|| = |c|||A||$.

c) $||A + B|| \leq ||A|| + ||B||$.

d) $||AB|| \leq ||A||||B||$.

A matrix norm on $M_{nn}$ is said to be *compatible* with the norm of a vector $||x||$ on $\mathbb{R}^n$ if

$$||Ax|| \leq ||A||||x||$$

holds true for all $n \times n$ matrices $A$ and for all $x \in \mathbb{R}^n$.

**Example 78.** The *Frobenius norm* $||A||_F$ of a matrix $A$ is obtained by taking the square root of the sum of the squares of the entries of $A$. Thus, if $A = [a_{ij}]$, then

$$||A||_F = \sqrt{\sum_{i,j=1}^{n} a_{i,j}^2}.$$

Suppose the given matrix $A$ is equal to

$$\begin{pmatrix} 3 & 1 \\ -2 & 5 \end{pmatrix}.$$

Then its Frobenius norm is $||A||_F = \sqrt{3^2 + 1^2 + (-2)^2 + 5^2} = \sqrt{39}$. Now observe that if $r_1 = (3, 1)$ and $r_2 = (-2, 5)$ are the row vectors of $A$, then the Euclidean norm of these vectors is $||r_1|| = \sqrt{3^2 + 1^2} = \sqrt{10}$ and $||r_2|| = \sqrt{(-2)^2 + 5^2} = \sqrt{29}$. So,

$$||A||_F = \sqrt{||r_1||^2 + ||r_2||^2}.$$

Similarly one can see if $c_1$ and $c_2$ are the column vectors of $A$, then

$$||A||_F = \sqrt{||c_1||^2 + ||c_2||^2}.$$

It is easy to see that these facts extends to $n \times n$ matrices in general.

**Proposition 28.** *The Frobenius norm is a matrix norm and it is compatible with the Euclidean norm.*

*Proof.* We first show that it is compatible with the Euclidean norm $|| \cdot ||_E$. Let us start by writing

$$A = \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{pmatrix},$$

where $A_i$ is the $i$th row of the matrix $A$. Then

$$\begin{aligned} ||Ax||_E &= \left|\left|\begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{pmatrix} x\right|\right|_E \\ &= \sqrt{||A_1x||_E^2 + \cdots + ||A_nx||_E^2} \\ &\leq \sqrt{||A_1||_E^2||x||_E^2 + \cdots + ||A_n||_E^2||x||_E^2} \\ &= \sqrt{||A_1||_E^2 + \cdots + ||A_n||_E^2} \cdot ||x||_E \\ &= ||A||_F||x||_E. \end{aligned}$$

Note that in the above we used the CSB inequality to conclude that $||A_ix||_E \leq ||A_i||_E||x||_E$ for $i = 1, \ldots, n$. Hence the Frobenius norm is compatible with the Euclidean norm. Let us now demonstrate that $||\cdot||_F$ is a matrix norm. Let $b_i$ denote the $i$th column of $B$. Using the representation for the product (matrix-column representation)

$$||AB||_F = ||[Ab_1 \cdots Ab_n]||_F,$$

we get:

$$\begin{aligned} ||AB||_F &= \sqrt{||Ab_1||_E^2 + \cdots + ||Ab_n||_E^2} \\ &\leq \sqrt{||A||_F^2||b_1||_E^2 + \cdots + ||A||_F^2||b_n||_E^2} \\ &= ||A||_F\sqrt{||b_1||_E^2 + \cdots + ||b_n||_E^2} \\ &= ||A||_F||B||_F. \end{aligned}$$

This proves property $d$) of the definition of a matrix norm. Properties $a$) through $c$) are also true, since the Frobenius norm is derived from the Euclidean norm. □

**Remark 43.** Unfortunately, there are situations where the Frobenius norm, also known as the Hilbert–Schmidt norm, is not convenient to use. In practice often the *operator norm* is used. To understand what we mean by operator norm, let us keep in mind that we want to measure the "size" of $A$ and compare it with $||x||$ and $||Ax||$; thus what we need is the inequality

$$||Ax|| \leq ||A||||x|| \text{ or } \frac{||Ax||}{||x||} \leq ||A|| \text{ for } x \neq 0.$$

The expression $\dfrac{||Ax||}{||x||}$ can be rewritten if we consider normalizing the vector $x$, which means we get a new vector $\overline{x}$ by dividing the vector $x$ by its length. This

is done by setting $\overline{x} = \dfrac{x}{||x||}$. Now, we only need to consider the vectors on the unit sphere as opposed to looking at all nonzero vectors $x$. Thus,

$$\frac{||Ax||}{||x||} = \frac{1}{||x||}||Ax|| = ||(\frac{x}{||x||})|| = ||A(\overline{x})||.$$

Furthermore,

$$\max_{x \neq 0} \frac{||Ax||}{||x||} = \max_{||\overline{x}||=1} ||A(\overline{x})||$$

always exists, and there is a particular vector $u$ for which $||Au||$ is maximal. The operator norm $||A||$ is then defined to be

$$||A|| = \max_{||x||=1} ||Ax||,$$

which defines a matrix norm on $M_{nn}$. The word *operator norm* reflects the fact that a matrix transformation arising from a square matrix is also called *linear operator*.

## Convergence in Normed Spaces

A sequence $\{x_n\}$ in a normed space is said to be *convergent* if for all $\epsilon > 0$ there exists a positive integer $N$ and an element $x \in X$ such that

$$||x_n - x|| < \epsilon \quad \text{whenever} \quad n > N.$$

We write this as $x_n \to x$.

**Theorem 80.** *If $X$ is a normed space, then the norm*

$$|| \cdot || : X \to \mathbb{R}$$

*is a continuous function on $X$.*

*Proof.* We need to show that $x_n \to x$ implies $||x_n|| \to ||x||$. This follows from the alternate triangle inequality

$$|||x_n|| - ||x||| \leq ||x_n - x||$$

and the fact that since $x_n \to x$, there exists an $N$ such that $||x_n - x|| < \epsilon$ for all $n > N$. □

**Theorem 81.** *The operation of addition and scalar multiplication in a normed space $X$ are jointly continuous.*

*Proof.* We will show if $x_n \to x$ and $y_n \to y$, then $x_n + y_n \to x + y$ and if $a_n \to a$ then $a_n x_n \to ax$. Since $x_n \to x$ and $y_n \to y$,

$$||(x_n + y_n) - (x + y)|| = ||(x_n - x) + (y_n - y)|| \leq ||x_n - x|| + ||y_n - y||$$

implies $x_n + y_n \to x + y$. Now

$$\begin{aligned} ||a_n x_n - ax|| &= ||a_n x_n - a_n x + a_n x - ax|| \\ &= ||a_n(x_n - x) + (a_n - a)x|| \\ &\leq |a_n|\,||x_n - x|| + |a_n - a|\,||x||, \end{aligned}$$

so since $\{a_n\}$ is bounded, $a_n \to a$ and $x_n \to x$, we obtain $a_n x_n \to ax$. □

The sequence $\{x_n\}$ is called a *Cauchy sequence* if, given $\epsilon > 0$, there exists a positive integer $N$ such that

$$||x_n - x_m|| < \epsilon \quad \text{whenever} \quad m, n > N.$$

When every Cauchy sequence in a normed space converges, then we say the normed space is *complete*. We have a special term for them:

**Definition 67.** A complete normed vector space is called a *Banach space*.

The spaces $\mathbb{R}^n$, $\mathbb{C}^n$, $\ell^2$, and $C[a, b]$ with the sup norm are all examples of Banach spaces. One can also prove that every finite dimensional normed vector space is a Banach space. All of the theorems we have seen about the convergence of sequences are valid in these spaces as well. For example the limit of a convergent sequence in a normed space is unique, and any convergent sequence in a normed space is a Cauchy sequence. Every normed space is a metric space; there is a metric induced by the norm and consequently a normed space may be identified as compact if and only if it is sequentially compact. Furthermore, in a vector space we can add vectors together so we have an idea of infinite series of real and complex numbers. Let $\{x_n\}$ be a sequence of elements in a normed linear space $X$; then we say $\displaystyle\sum_{k=1}^{\infty} x_k$ converges if the sequence of partial sums $s_n = \displaystyle\sum_{k=1}^{n} x_k$ converges.

**Definition 68.** A series $\displaystyle\sum_{k=1}^{\infty} x_k$ *absolutely convergent* if the series $\displaystyle\sum_{k=1}^{\infty} ||x_n||$ (of real numbers) is convergent.

This is just a generalization to Banach spaces of the definition of absolute convergence of sequence of real numbers. If a series is absolutely convergent in a normed linear space $X$, it need not converge in $X$. This is illustrated in the following example:

**Example 79.** Consider the set $\mathcal{P}[0, 1]$ of all polynomials with the supremum norm. In this normed linear space, consider the series $\displaystyle\sum_{n=0}^{\infty} \frac{x^n}{n!}$ for $x \in [0, 1]$. The series

$$\sum_{n=0}^{\infty} \frac{||x||^n}{n!} \leq \sum_{n=0}^{\infty} \frac{1}{n!}$$

converges to a constant less than $e$. But the constants can be considered as polynomials, thus the series $\sum_{n=0}^{\infty} \frac{x^n}{n!}$ converges absolutely. However, if we consider the sequence of partial sums $s_n = \sum_{k=0}^{n} \frac{x^k}{k!}$, we see that $s_n \to e^x$ as $n \to \infty$ in the norm. Since $e^x \notin \mathcal{P}[0,1]$, the series is not convergent in $\mathcal{P}[0,1]$.

The following theorem shows that absolute convergence implies convergence in Banach spaces.

**Theorem 82.** *A normed vector space $X$ is a Banach space if and only if every absolutely convergent series in $X$ is convergent.*

*Proof.* Suppose $X$ is a Banach space and that $\sum_{k=1}^{\infty} x_k$ is an absolutely convergent series in $X$. Then by definition $\sum_{k=1}^{\infty} ||x_k||$ converges. Using the triangle inequality we get

$$||s_n - s_m|| = || \sum_{k=m}^{n} x_k|| \leq \sum_{k=m}^{n} ||x_k|| < \epsilon$$

for all sufficiently large $m, n$ with $m < n$. To prove the converse, assume every absolutely convergent series to be convergent and let $\{x_n\}$ be a Cauchy sequence in $X$. Thus, for every $\epsilon > 0$ we can find a positive integer $N$ so that $||x_n - x_m|| < \epsilon$ when $m, n > N$. In particular we may take $\epsilon = \frac{1}{2^k}$ with $k = 1, 2 \ldots$. For each positive integer $k$, there exists a positive integer $N_k$ such that

$$||x_n - x_m|| < \frac{1}{2^k} \quad \text{for all} \quad m, n > N_k.$$

From this we get,

$$||x_{n_{k+1}} - x_{n_k}|| < \frac{1}{2^k} \quad \text{for} \quad k = 1, 2, \ldots$$

It follows that

$$\sum_{k=1}^{\infty} ||x_{n_{k+1}} - x_{n_k}|| \leq \sum_{k=1}^{\infty} \frac{1}{2^k} = 1,$$

so the series $\sum_{k=1}^{\infty} ||x_{n_{k+1}} - x_{n_k}||$ is convergent. By hypothesis the series $\sum_{k=1}^{\infty} (x_{n_{k+1}} - x_{n_k})$ is convergent. This implies that $\{x_{n_k}\}$ converges. Hence the Cauchy sequence $\{x_n\}$ has a convergent subsequence $\{x_{n_k}\}$, but by the property of Cauchy sequences this means the Cauchy sequence $\{x_n\}$ converges. Hence $X$ is a complete normed space. □

## Exercises

1. Show that $d(x, y) = |e^x - e^y|$ is a metric on $\mathbb{R}$.

2. Let $S$ denote the collection of all real sequences $x = \{x_n\}$. Show that the expression

$$d(x, y) = \sum_{i=1}^{\infty} \left(\frac{1}{2^i}\right) \frac{|x_i - y_i|}{1 + |x_i - y_i|}$$

defines a metric on $S$.

3. If $d$ is a metric on $M$, show that

$$|d(x, z) - d(y, z)| \leq d(x, y),$$

for any $x, y, z \in M$.

4. Let $f : [0, 1] \to \mathbb{R}$ be continuously differentiable, with $f(0) = 0$. Prove that

$$\sup_{0 \leq x \leq 1} |f(x)| \leq \sqrt{\int_0^1 (f'(x))^2 \, dx}.$$

Hint: Apply CSB inequality to $|f(x)| = |\int_0^x f'(t)dt|$.

5. Prove that in an inner product space we have additivity in the second slot as well as the first slot.

6. Prove that in an inner product space we have conjugate homogeneity in the second slot, i.e., prove that

$$\langle u, \alpha v \rangle = \overline{\alpha} \langle u, v \rangle,$$

for all scalars $\alpha$.

7. Let $u$ and $v$ be vectors in an inner product space $V$. Give an alternate proof of the CSB inequality by answering the following questions:

a) Let $u \neq 0$ and $t$ be a scalar. Then consider

$$\langle tu + v, tu + b \rangle \geq 0,$$

for all values of $t$. Expand this inequality to obtain a quadratic inequality of the form

$$at^2 + bt + c \geq 0.$$

What are $a, b$, and $c$ in terms of $u$ and $v$?

b) Use your knowledge of quadratic equations and their graphs to obtain a condition on $a, b$, and $c$ for which the inequality in part a) is true.

c) Show that in terms of $u$ and $v$, your condition in part b) is equivalent to the CSB inequality.

8. Show that

$$||x||_\infty \leq ||x||_2 \leq ||x||_1,$$

for any $x \in \mathbb{R}^n$. Also show the following:

$$\begin{aligned} ||x||_1 &\leq n||x||_\infty, \\ ||x||_1 &\leq \sqrt{n}||x||_2. \end{aligned}$$

9. Show that if $A = \{a_{ij}\}$ is an $m \times n$ matrix then

$$||A|| = \max_{1 \leq i \leq n} \left( \sum_{j=1}^{m} |a_{ij}|^2 \right)^{1/2}$$

is a norm on the vector space $M_{m \times n}$ of all $m \times n$ matrices.

10. Let $A$ and $B$ be $n \times n$ complex matrices

a) Show that $\langle A, B \rangle = \text{tr}(AB^*)$ defines an inner product on the space of $n \times n$ complex matrices.

b) Prove that $|\text{tr}(AB^*)|^2 \leq \text{tr}(AA^*)\text{tr}(BB^*)$.

11. In a normed space, prove that

a) $||x - y|| \geq |||x|| - ||y|||$,

b) $||(1/\lambda)x|| = 1$ if $\lambda = ||x||$, $x \neq 0$.

12. Let $\{x_n\}$ be a sequence in a normed space and suppose that $\sum_{k=1}^{\infty}(x_k - x_{k+1})$ is absolutely convergent. Determine whether $\{x_n\}$ is Cauchy or convergent.

13. Let $H$ be an inner product space with

$$\langle f, g \rangle = \frac{1}{2\pi} \int_0^{2\pi} f(t)\overline{g(t)}dt.$$

a) Show that

$$S = \{f_n(t) = e^{int} : n \text{ is an integer}\}$$

is on orthonormal set.
(Recall that $e^{int} = \cos nt + i \sin nt$, $0 \leq t \leq 2\pi$ and $\overline{e^{int}} = e^{-int}$)

b) Let $W = \text{span}(S)$ and let $T$ and $U$ be operators defined on $W$ by

$$T(f_n) = f_{n+1} \text{ and } U(f_n) = f_{n-1}.$$

Show that $U = T^*$. (Recall that $U = T^*$ if $\langle Tf_n, f_m\rangle = \langle f_n, Uf_m\rangle$).

14. Suppose $C[-1,1]$ is the vector space of real-valued continuous functions on $[-1,1]$ with the inner product giving by

$$\langle f, g\rangle = \int_{-1}^{1} f(x)g(x)\,dx$$

for $f, g \in C[-1,1]$. Let $\phi$ be the linear functional on $C[-1,1]$ defined by $\phi(f) = f(0)$. Show that $\phi(f) = f(0)$. Show that there does **not** exist $g \in C[-1,1]$such that

$$\phi(f) = \langle f, g\rangle$$

for every $f \in C[-1,1]$. Is this a contradiction to Riesz representation theorem?

15. The *diameter* of a set of a nonempty set $A$ of a metric space $(M,d)$ is defined by

$$\text{diam}(A) = \sup\{d(a,b) : a, b \in A\}.$$

Show that $A$ is bounded if and only if $\text{diam}(A)$ is finite.

## 2.2 Fixed Point Theorems and Applications

One of the most common problems in mathematics is the following:

Given a linear mapping $T$ and the image $y$, solve $Tx = y$ for $x$.

If the mapping $T : \mathbb{R}^n \to \mathbb{R}^n$, then this problem becomes that of solving a set of simultaneous linear equations. We often encounter the need to solve $f(x) = 0$, where $f$ is a real-valued function. Of course when $f$ is a linear or a quadratic function, we know how to solve it. But for most other functions we use some method of approximating the roots. For example *Newton's method* provides one such approximation under certain conditions. We assume $x_0$ to be an approximation to the root and then calculate $x_1 = x_0 - \dfrac{f(x_0)}{f'(x_0)}$; next we will use $x_1$ in this formula to obtain a better approximation $x_2$ and so on. Such a process is called *iterative*. Our hope is to get a sequence of points $x_0, x_1, \ldots$ such that this sequence converges to the root of $f(x) = 0$.

Completeness is a useful property if one is interested in solving equations. For example how does one compute $\sqrt{2}$ by hand? You would most likely start by finding an approximate solution to the equation $x^2 = 2$ and then seek ways to improve the estimate. Better and better estimates give you a sequence of points,

but how do we know this sequence converges? The completeness of $\mathbb{R}$ comes to the rescue, since the sequence converges to a point in $\mathbb{R}$. The modern metric space version of the method of successive approximation was first explicitly stated in Banach's thesis in 1922. He considered a complete metric space $M$ and a contraction mapping $f : M \to M$ and showed how to get the unique solution to $f(x) = x$. This theorem is now referred as the Banach Contraction Mapping Theorem which we will explain in detail below. However, we first examine a continuous function $f : [-1, 1] \to [-1, 1]$ shown in Figure 2.9.

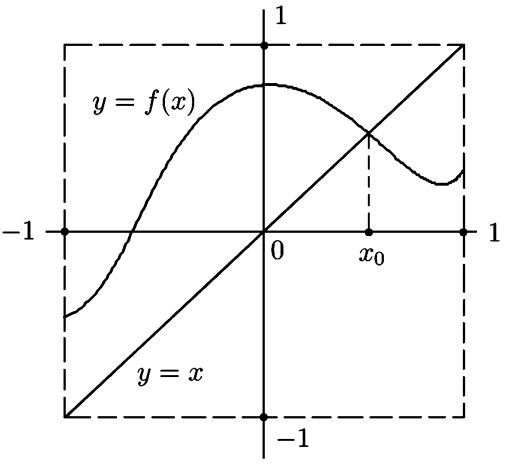



Figure 2.9: A continuous function $f : [-1, 1] \to [-1, 1]$ and its fixed point $x_0$

Figure 2.9 indicates that the graph of $f$ must cross the diagonal line; more precisely, there must be a point $x_0 \in [-1, 1]$ such that $f(x_0) = x_0$. This is easy to see if we first set up the function

$$F(x) = f(x) - x.$$

Observe that $F$ is defined on $[-1, 1]$ and $F$, being the difference of two continuous functions, is continuous on $[-1, 1]$. Furthermore, $F(1) \geq 0$ and $F(-1) \leq 0$. Therefore, we can apply the Intermediate Value Theorem to conclude that there exists a point $x_0$ in $[-1, 1]$ such that $F(x_0) = 0$ or equivalently $f(x_0) = x_0$.

**Definition 69** (Fixed Point)**.** A **fixed point** of a mapping $f : A \to A$ is an $x \in A$ which is mapped onto itself (is "kept fixed" by $f$), that is,

$$f(x) = x;$$

the image $f(x)$ coincides with $x$.

Given a mapping there is no guarantee that it has a fixed point, let alone that it has a unique fixed point. For example, a translation $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x + a$ where $a \neq 0$, has no fixed points. On the other hand, $f : \mathbb{R} \to \mathbb{R}$ where $f(x) = x^2$

has two fixed points 0 and 1 as we can see $f(0) = 0$ and $f(1) = 1$. The mapping (projection) $P$ of $\mathbb{R}^2$ onto the $x$-axis where $P(x, y) = (x, 0)$ has infinitely many fixed points, namely the entire $x$-axis where points are of the form $(x, 0)$.

**Definition 70** (Contraction)**.** Let $X = (X, d)$ be a metric space. A mapping $T : X \to X$ is called a **contraction on** $X$ if there is a positive constant $\alpha < 1$ such that for all $x, y \in X$

$$d(Tx, Ty) \leq \alpha \; d(x, y).$$

Geometrically this implies that the images of any two points $x, y$ are closer together than the points themselves.

**Observation 6.** Let $I = [a, b]$ or $[a, b)$ for $b < \infty$ and $f : I \to I$. If $\alpha = \sup_{x \in I} |f'(x)| < 1$, then $f$ is a contraction on $I$. To see this observe that for any $x, y \in I$, by the Mean Value Theorem, we have for some $\zeta$ between $x$ and $y$

$$f(x) - f(y) = f'(\zeta)(x - y).$$

Since we know that $\alpha = \sup_{x \in I} |f'(x)| < 1$, we have

$$|f(x) - f(y)| = |f'(\zeta)||x - y| \leq \alpha |x - y|.$$

## Banach Fixed Point Theorem

The following theorem called the Banach Fixed Point Theorem, Banach Contraction Mapping Theorem, or simply the Contraction Mapping Theorem is one of the jewels of analysis. It is an existence and uniqueness theorem for fixed points for a class of mappings, called contractions. The strength of it comes from the fact that it has a constructive proof. In other words, it gives a **constructive** procedure for obtaining better and better approximations to the fixed point. This procedure, as we mentioned before, is called an iteration process. Iteration procedures are used in nearly every branch of applied mathematics, and convergence proofs and error estimates are very often obtained by an application of Banach's Fixed Point Theorem. In the following we also present some fundamental applications of this powerful theorem.

**Theorem 83** (Banach Fixed Point Theorem or Contraction Mapping Theorem)**.** *Let $X = (X, d)$ be a complete metric space. Then any contraction $f : X \to X$ has precisely one fixed point.*

*Proof.* We construct a sequence $\{x_n\}$ and show that it is Cauchy, so that it converges in the complete space $X$, and then we prove that its limit $x$ is a fixed point of $f$ and $f$ has no further fixed points. We will do this proof in steps.

a) Choose an arbitrary point $x_0 \in X$. Define a sequence $\{x_n\}$ inductively as follows:

$$x_{n+1} = f(x_n).$$

So we have

$$\begin{aligned} x_1 &= f(x_0) \\ x_2 &= f(x_1) = f(f(x_0)) = f^2(x_0) \\ &\vdots \\ x_n &= f^n(x_0) \\ &\vdots \end{aligned}$$

where $f^n$ is the $n$th composition of $f$.

b) We will show that this sequence is Cauchy. Observe that

$$\begin{aligned} d(x_{n+1}, x_n = d(f(x_n), f(x_{n-1})) &\leq kd(x_n, x_{n-1}) = kd(f(x_{n-1}), f(x_{n-2})) \\ &\leq k^2 d(x_{n-1}, x_{n-2}) \leq \cdots \leq k^n d(x_1, x_0), \end{aligned}$$

where we have used the fact that $f$ is a contraction with contraction constant $k < 1$. Thus if $n > m$, we have that

$$\begin{aligned} d(x_m, x_n) &\leq d(x_m, x_{m+1}) + d(x_{m+1}, x_{m+2}) + \cdots + d(x_{n-1}, x_n) \\ &\leq (k^m + k^{m+1} + \cdots + k^{n-1}) d(x_1, x_0) \\ &= [k^m(1 + k + \cdots + k^{n-m-1})]\, d(x_1, x_0). \end{aligned}$$

We now use a well-known identity for $r < 1$,

$$1 + r + r^2 + \cdots + r^n = \frac{1 - r^{n+1}}{1 - r},$$

to conclude

$$k^m(1 + k + \cdots + k^{n-m-1}) = k^m \frac{1 - k^{n-m}}{1 - k}.$$

Since we know that $n > m$ we have

$$0 < k < 1 \implies 1 - k^{n-m} < 1.$$

So we have

$$d(x_m, x_n) < \frac{k^m}{1 - k} d(x_0, x_1).$$

Then it is clear that $d(x_m, x_n) \to 0$ as $n, m \to \infty$. Since $\{x_n\}$ is Cauchy in a complete metric space, $x_n \to x^* \in X$.

c) Since $f$ is continuous (why?), we know from the sequential characterization of continuity that $f(x_n) \to f(x^*)$. So we have

$$f(x^*) = \lim_{n\to\infty} f(x_n) = \lim_{n\to\infty} x_{n+1} = x^* \implies f(x^*) = x^*.$$

Thus we have found a fixed point.

d) Now we will prove uniqueness. Suppose there is another fixed point $y$ such that $f(y) = y$. Observe

$$d(y, x^*) = d(f(y), f(x^*)) \leq k\, d(y, x^*),$$

$$\Rightarrow (1-k)d(y, x^*) \leq 0$$

since we have a contraction, we must have $d(y, x^*) = 0$, therefore $y = x^*$. □

**Remark 44** (Error Estimate)**.** In the proof of the Banach Fixed Point Theorem we are given the method to construct a fixed point as the limit of a sequence $\{x_n\}$. If we take $n \to \infty$ we have the estimate for the error

$$d(x_m, x) \leq \frac{k^m}{1-k} d(x_1, x_0),$$

where $k < 1$ is the contraction constant.

## Applications

In the following several applications of the Banach Fixed Point Theorem are given.

## Applications to Solutions of the Equation $f(x) = x$

**1.** Finding a root of the equation $e^{-x} = x$.
We see that a root of $f(x) = e^{-x}$ is in $[\frac{1}{2}, 1]$ because for $x \in [\frac{1}{2}, 1]$, we have $|f'(x)| = |-e^{-x}| = e^{-x} \leq e^{-\frac{1}{2}} = 0.6065 < 1$. This implies that $f$ is a contraction on $[\frac{1}{2}, 1]$ and that the iteration $x_n = f(x_{n-1})$ has a limit. So we will solve this by employing an iterative scheme. Let us start with $x_0 = 0.5$; then we have that

$$\begin{aligned} x_1 &= f(x_0) = e^{-0.5} = 0.6065 \\ x_2 &= f(x_1) = e^{-0.6065} = 0.5452 \\ &\vdots \end{aligned}$$

Continuing this iteration scheme we obtain,

| n | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| $x_n$ | 0.5000 | 0.6065 | 0.5452 | 0.5797 | 0.5601 | 0.5712 | 0.5649 | 0.5684 |

| n | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|
| $x_n$ | 0.5664 | 0.5676 | 0.5669 | 0.5673 | 0.5671 | 0.5672 | 0.5671 | 0.5672 |

Note that, in the above example, we have $x = 0.567$ is a fixed point correct to three decimal places. Starting with some arbitrary point $x_0$, we can explore this visually as shown by the following Figure 2.10 in which the iterative sequence approaches the fixed point that is the solution to the equation.

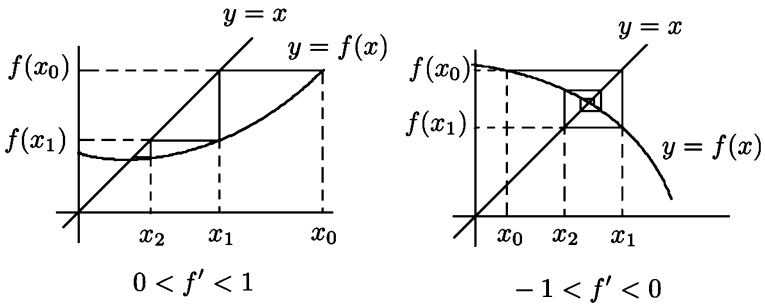



Figure 2.10: Iterations

**2.** Solve the equation $x^3 + x - 1 = 0$.
We let $F(x) = x^3 + x - 1$ and observe that

$$\begin{aligned} F\left(\frac{1}{2}\right) &= -\frac{3}{8} < 0 \\ F\left(\frac{3}{4}\right) &= \frac{11}{64} > 0, \end{aligned}$$

which implies that the root is in $[\frac{1}{2}, \frac{3}{4}]$. We have some possibilities to solve for $x$:

$$x = 1 - x^3 = f(x), \tag{2.1}$$

$$x = \frac{1}{1+x^2} = g(x), \tag{2.2}$$

$$x = \frac{\sqrt{x}}{\sqrt{1+x^2}} = h(x). \tag{2.3}$$

So, we will check the derivatives of each of these functions. For (2.1),

$$|f'(x)| = |-3x^2| = 3x^2 > 1.$$

This is close to the solution and the iteration sequence $\{x_n\}$ will not be convergent. Now we take (2.2),

$$|g'(x)| = \left|-\frac{2x}{(1+x^2)^2}\right| = \frac{2x}{(1+x^2)^2} \leq \frac{2\cdot\frac{3}{4}}{(1+\frac{1}{4})^2} = \frac{24}{25} < 1.$$

This means that $x_n = g(x_{n-1})$ is convergent. Now we look at (2.3), notice that $|h'(x)| < \frac{4}{9}$. This implies that $x_n = h(x_{n-1})$ converges even faster. Note that we have special methods such as Newton's method, the square root method, etc. to solve equations. Thus, in general if we have $f : [a,b] \to [a,b]$ and $\alpha$ such that $\alpha = \sup_{x\in[a,b]} |f'(x)| < 1$, then the equation $f(x) = x$ has a unique solution $\overline{x} \in [a,b]$ which can be approximated by iterations $(x_0, x_1, \dots)$ where $x_{n+1} = f(x_n)$, and with the error estimate given by

$$|x_{n+1} - \overline{x}| \leq \frac{\alpha}{1-\alpha}\,|x_{n+1} - x_n| \leq \frac{\alpha^n}{1-\alpha}|x_1 - x_0|.$$

## Applications to Systems of Algebraic Linear Equations

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be given by

$$\vec{y} = A\vec{x} + \vec{b},$$

where $A = (a_{ij})$ and $B = (b_i)$ are fixed. We can write this as

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}. \tag{2.4}$$

We write this in a more compact form as

$$y_i = \sum_{j=1}^{n} a_{ij}x_j + b_i, \quad i = 1, 2, \dots, n.$$

To solve the equation $\vec{y} = A\vec{x} + \vec{b}$ apply the Banach Fixed Point Theorem; but for this theorem, we need a complete metric space and a contraction on it. We already know $\mathbb{R}^n$ is complete metric space, so whether the mapping

$$y = Tx = Ax + b$$

is a contraction on $(\mathbb{R}^n, d)$ depends on the metric $d$ as illustrated in the following.

a) Suppose for $(\mathbb{R}^n, d)$ the metric $d(\cdot,\cdot)$ given by

$$d(x,z) = \max_{1\leq i\leq n} |x_i - z_i|.$$

We have that

$$\begin{aligned} d(y', y'') &= \max_i |y_i' - y_i''| = \max_i \left| \sum_{j=1}^n a_{ij} x_j' + b_i - \left( \sum_{j=1}^n a_{ij} x_j'' + b_i \right) \right| \\ &= \max_i \left| \sum_{j=1}^n a_{ij} (x_j' - x_j'') \right| \leq \max_i \sum_{j=1}^n |a_{ij}||x_j' - x_j''| \\ &\leq \max_i \sum_{j=1}^n |a_{ij}| \max_j |x_j' - x_j''| = \max_i \sum_{j=1}^n |a_{ij}| d(x', x''). \end{aligned}$$

Thus, using the Banach Fixed Point Theorem we arrive at the following theorem:

**Theorem 84** (Row Sum Criterion)**.** *If a system*

$$x = Ax + b \quad (A,\ b \ \text{ given})$$

*of $n$ linear equations in $n$ unknowns, $(x_1, x_2, \ldots, x_n)$, satisfies*

$$\alpha = \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}| < 1, \tag{2.5}$$

*then there is precisely one solution $x$. This solution can be obtained as the limit of the iterative sequence $(x^{(0)}, x^{(1)}, \ldots)$, where $x^{(0)}$ is arbitrary and*

$$x^{(m+1)} = Ax^{(m)} + b, \quad m = 0, 1, 2, \ldots$$

*with error bounds of*

$$d(x^{(m)}, x) \leq \frac{\alpha^m}{1 - \alpha} d(x^{(0)}, x^{(1)}).$$

Now we will work through a concrete example. Consider the following set of algebraic equations,

$$\begin{cases} \frac{4}{5} x_1 - \frac{1}{2} x_2 - \frac{1}{4} x_3 = -1 \\ -\frac{1}{3} x_1 + \frac{4}{3} x_2 + \frac{1}{4} x_3 = 2 \\ -\frac{1}{4} x_1 - \frac{2}{15} x_2 + \frac{3}{4} x_3 = 3. \end{cases}$$

To apply the Row Sum Criterion, we first rewrite the equations in the form $x = Ax + b$; we have

$$\begin{cases} x_1 = \frac{1}{5} x_1 + \frac{1}{2} x_2 + \frac{1}{4} x_3 - 1 \\ x_2 = \frac{1}{3} x_1 - \frac{1}{3} x_2 - \frac{1}{4} x_3 + 2 \\ x_3 = \frac{1}{4} x_1 + \frac{2}{15} x_2 + \frac{1}{4} x_3 + 3 \end{cases}$$

which can also be written as

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} \frac{1}{5} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{3} & -\frac{1}{3} & -\frac{1}{4} \\ \frac{1}{4} & \frac{2}{15} & \frac{1}{4} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} -1 \\ 2 \\ 3 \end{pmatrix}.$$

We also know that

$$\begin{aligned} \alpha &= \max_{1\le i\le 3} \sum_{j=1}^{3} |a_{ij}| = \max\left\{ \frac{1}{5}+\frac{1}{2}+\frac{1}{4}, \frac{1}{3}+\frac{1}{3}+\frac{1}{4}, \frac{1}{4}+\frac{2}{15}+\frac{1}{4} \right\} \\ &= \max\left\{ \frac{19}{20}, \frac{11}{12}, \frac{19}{30} \right\} = \frac{19}{20} < 1. \end{aligned}$$

So we can begin with an iteration scheme, $x^{(0)} = (0,0,0)$, and we can iterate to find

$$\begin{aligned} x^{(1)} &= (-1,2,3), \\ x^{(5)} &= (0.507535, 0.942640, 4.303294), \\ x^{(10)} &= (0.641658, 0.835535, 4.360547), \\ x^{(19)} &= (0.639519, 0.841889, 4.362843), \\ x^{(20)} &= (0.639559, 0.841833, 4.362843). \end{aligned}$$

This leads to the exact solution of

$$x = (0.639548, 0.841853, 4.362845).$$

b) Suppose in $(\mathbb{R}^n, d)$ the metric $d(\cdot,\cdot)$ is given by

$$d(x,z) = \sum_{i=1}^{n} |x_i - z_i|. \tag{2.6}$$

Using this metric, in the following we obtain a condition (the column sum condition) under which $T$ is a contraction.

$$\begin{aligned} d(y',y'') &= \sum_{i=1}^{n} |y_i' - y_i''| = \sum_{i=1}^{n} \left| \sum_{j=1}^{n} a_{ij}x_j' + b_i - (\sum_{j=1}^{n} a_{ij}x_j'' + b_i) \right| \\ &= \sum_{i=1}^{n} \left| \sum_{j=1}^{n} a_{ij}(x_j' - x_j'') \right| \le \sum_{i=1}^{n}\sum_{j=1}^{n} |a_{ij}||x_j' - x_j''| \\ &\le \max_{1\le j\le n} \sum_{i=1}^{n} |a_{ij}| \sum_{j=1}^{n} |x_j' - x_j''| \\ &= \max_{1\le j\le n} \sum_{i=1}^{n} |a_{ij}|\, d(x,x''). \end{aligned}$$

**Theorem 85** (Column Sum Criterion)**.** *If*

$$\beta = \max_{1\le j\le n} \sum_{i=1}^{n} |a_{ij}| < 1, \tag{2.7}$$

*then $T$ is a contraction and the system $x = Ax + b$ has precisely one solution.*

By the same concrete exercise as before we have

$$\begin{aligned} \beta = \max_{j} \sum_{i=1}^{3} |a_{ij}| &= \max\left\{ \frac{1}{5}+\frac{1}{3}+\frac{1}{4}, \quad \frac{1}{2}+\frac{1}{3}+\frac{2}{15}, \quad \frac{1}{4}+\frac{1}{4}+\frac{1}{4} \right\} \\ &= \frac{29}{30} < 1. \end{aligned}$$

c) If we use the Euclidean metric on $\mathbb{R}^n$

$$d(x,z) = \left( \sum_{i=1}^{n} |x_i - z_i|^2] \right)^{\frac{1}{2}},$$

we then see

$$\begin{aligned} d(y',y'')^2 &= \sum_{i=1}^{n} |y_i' - y_i''|^2 = \sum_{i=1}^{n} \left| \sum_{j=1}^{n} (a_{ij}x_j' + b_i) - \sum_{j=1}^{n} (a_{ij}x_j'' - b_i) \right|^2 \\ &= \sum_{i=1}^{n} \left| \sum_{j=1}^{n} a_{ij}(x_j' - x_j'') \right|^2 \le \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2 \sum_{j=1}^{n} (x_j' - x_j'')^2 \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2 d(x',x'')^2. \end{aligned}$$

So we have if

$$\gamma = \left( \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2 \right)^{\frac{1}{2}} < 1,$$

then $T$ is a contraction and the system has precisely one solution.
In our exercise we obtain that

$$\begin{aligned} \gamma &= \left[ \left( \frac{1}{25}+\frac{1}{4}+\frac{1}{16} \right) + \left( \frac{1}{9}+\frac{1}{9}+\frac{1}{16} \right) + \left( \frac{1}{16}+\frac{4}{225}+\frac{1}{16} \right) \right]^{\frac{1}{2}} \\ &= \left( \frac{39}{50} \right)^{\frac{1}{2}} < 1. \end{aligned}$$

Note the iterative sequence turns out to converge faster than in part a).

## Applications to Differential Equations

Consider the Cauchy initial value problem for ordinary differential equations of the first order

$$\begin{cases} x' = f(t,x) \\ x(t_0) = x_0 \end{cases} . \tag{2.8}$$

We examine the solution of this initial value problem with the help of the Contraction Mapping Theorem.

**Definition 71** (Lipschitz Condition)**.** We say $f$ satisfies a **Lipschitz Condition** if there exists $L > 0$ such that

$$|f(x) - f(y)| \leq L|x - y|.$$

**Theorem 86** (Picard Existence and Uniqueness Theorem, 1890)**.** *Let $f$ be continuous on the rectangle,*

$$R = \{(t,x) : |t - t_0| \leq a, |x - x_0| \leq b\} = [t_0 - a, t_0 + a] \times [x_0 - b, x_0 + b].$$

*Suppose that $f$ satisfies a Lipschitz condition with respect to the second argument which implies that there exists $L > 0$ such that*

$$|f(t,x_1) - f(t,x_2)| \leq L|x_1 - x_2| \quad \forall (t,x_1),(t,x_2) \in \mathbb{R}^2.$$

*Then the Cauchy initial value problem (2.8) has a unique solution (Figure 2.11). This solution exists on an interval $[t_0 - \beta, t_0 + \beta]$ where*

$$\beta < \min\left\{a, \frac{b}{M}, \frac{1}{L}\right\}, \quad M = \sup_{(t,x)\in\mathbb{R}} |f(t,x)|.$$

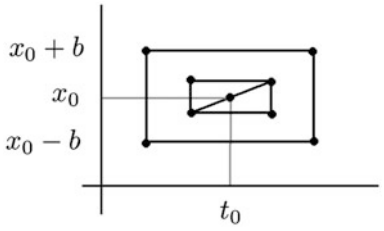



Figure 2.11: Rectangle in Picard's theorem

The solution passes through $(t_0, x_0)$.

*Proof.* We will proceed in steps.

a) Let $X$ be the Banach space $C(J)$, $J = [t_0 - \beta, t_0 + \beta]$, with the sup norm and

$$\tilde{C} = \{x \in C(J) : x_0 - M\beta \leq x(t) \leq x_0 + M\beta \quad \forall t \in J\},$$

where $M\beta < b$ . $\tilde{C}$ is a closed subspace which implies that $\tilde{C}$ is complete.

b) Consider the initial value problem given in 2.8; we form the following integral equation

$$x(t) = x(t_0) + \int_{t_0}^{t} f(s, x(s))ds \quad \text{and} \quad Tx(t) = x_0 + \int_{t_0}^{t} f(s, x(s))ds.$$

c) Consider $T : \tilde{C} \to \tilde{C}$, and notice that

$$\begin{aligned} |Tx(t) - x_0| &= \left| \int_{t_0}^{t} f(s, x(s))ds \right| \\ &\leq \int_{t_0}^{t} |f(s, x(s))|ds \leq M|t - t_0| \leq M\beta. \end{aligned}$$

This implies that $Tx(t) \in \tilde{C}$.

d) Now we will show that $T$ is a contraction on $\tilde{C}$ using the hypothesis that $T$ is Lipschitz in the second variable:

$$\begin{aligned} |Tx(t) - Ty(t)| &= \left| \int_{t_0}^{t} [f(s, x(s)) - f(s, y(s))]ds \right| \\ &\leq \int_{t_0}^{t} |f(s, x(s)) - f(s, y(s))|ds \leq L \int_{t_0}^{t} |x(s) - y(s)|ds \\ &\leq L||x - y||_\infty |t - t_0| \leq L\beta||x - y||_\infty \quad \forall t \in J. \end{aligned}$$

This implies that

$$||Tx - Ty||_\infty \leq L\beta||x - y||_\infty.$$

e) The Banach Fixed Point Theorem implies $T$ has a unique fixed point $x \in \tilde{C}$, that is, a continuous function $x$ on $J$ satisfying $x = Tx$ which implies

$$x(t) = x_0 + \int_{t_0}^{t} f(s, x(s))ds,$$

and if $(s, x(s)) \in \mathbb{R}^2$ and $f$ is continuous we may differentiate and this satisfies the Cauchy initial value problem (2.8). □

**Corollary 14.** *The Contraction Mapping Theorem also implies that the unique solution of (2.8) is the limit of the sequence, $\{x_0, x_1, x_2, \dots\}$ obtained by the Picard iteration*

$$x_{n+1}(t) = x_0 + \int_{t_0}^{t} f(s, x_n(s))ds.$$

Note that a bigger interval than $[t_0 - \beta, t_0 + \beta]$ can be obtained by using a different metric on $C(J)$. Also, with such a metric we can solve the Cauchy problem in the strip

$$S = \{(t, x) : |t - t_0| \leq a,\ x \in \mathbb{R}\}.$$

## Applications to Integral Equations

A little more than a hundred years ago, Fredholm made a study of integral equations of the type

$$x(t) = \mu \int_a^b k(t,s)x(s)\,ds + v(t).$$

These arose from questions of theoretical physics and the aim was to solve the function $x$ in terms of the function $v$. Since one can think of an integral as a limit of a finite sums, one can view the Fredholm integral equation as an infinite dimensional counterpart of finite dimensional linear systems. There is also a close connection between differential and integral equations and some problems may be formulated either way. In this section, we examine an application of the Contraction Mapping Theorem to integral equations. We can consider integral equations in various function spaces and we will consider them here in $C[a,b]$, the space of continuous functions on $[a,b]$. We start by defining the Fredholm integral equation.

**Definition 72.** The Fredholm Integral equation of the second kind is an equation of the form

$$x(t) = \mu \int_a^b k(t,s)x(s)ds + v(t), \tag{2.9}$$

where $k, v$ are given functions, $\mu \in \mathbb{R}$ is a parameter, and $x$ is an unknown function on $[a,b]$, $k(x,y) \in C([a,b] \times [a,b])$ is called the *kernel* of this integral equation.

Note that a Fredholm equation of the first kind is given by

$$\int_a^b k(t,s)x(s)ds = x(t).$$

**Theorem 87** (Fredholm Integral Equation)**.** *Suppose* $k \in C([a,b] \times [a,b])$, $v \in C[a,b]$ *and*

$$|\mu| < \frac{1}{M(b-a)}, \quad M = \sup_{s,t\in[a,b]} |k(t,s)|.$$

*Then* (2.9) *has a unique solution* $x \in C[a,b]$*. This function* $x$ *is the limit of the iterative sequence* $(x_0, x_1, \dots)$*, where* $x_0$ *is any function* $C[a,b]$ *and*

$$x_{n+1}(t) = v(t) + \mu \int_a^b k(t,s)x_n(s)ds, \quad n = 0, 1, \dots$$

*Proof.* We can write equation (2.9) as $x = Tx$ where

$$Tx(t) = v(t) + \mu \int_a^b k(t,s)x(s)ds.$$

The proof then follows from the Banach Fixed Point Theorem if we prove that $T$ is a contraction on $C[a,b]$. We have

$$\begin{aligned} d(Tx,Ty) &= \sup_{t\in[a,b]} |Tx(t) - Ty(t)| \\ &= \sup_{t\in[a,b]} \left| v(t) + \mu \int_a^b k(t,s)x(s)ds - \left[ v(t) + \mu \int_a^b k(t,s)y(s)ds \right] \right| \\ &= \sup_{t\in[a,b]} \left| \mu \int_a^b k(t,s)[x(s) - y(s)]ds \right| \\ &\leq |\mu| \sup_{t\in[a,b]} \int_a^b |k(t,s)|x(s) - y(s)|ds \\ &\leq |\mu| M(b-a)d(x,y) = \alpha d(x,y). \end{aligned}$$

Since

$$\alpha = |\mu| M(b-a) < 1,$$

we have that $T$ is a contraction on $C[a,b]$. □

**Remark 45.** We can also apply the above method to the nonlinear integral equation

$$x(t) - \mu \int_a^b k(t,s,x(s))ds = v(t),$$

when

$$|k(t,s,u_1) - k(t,s,u_2)| \leq L|u_1 - u_2|,$$

and

$$|\mu| < \frac{1}{L(b-a)}.$$

Now we will look at the Volterra integral equation.

**Definition 73** (Volterra Integral Equation)**.** Suppose $v \in C[a,b]$ and $k \in C([a,b] \times [a,t])$. Then the *Volterra integral equation* is an equation of the form

$$x(t) = \mu \int_a^t k(t,s)x(s)ds + v(t). \tag{2.10}$$

Note that the difference between the Volterra integral equation above and Fredholm integral equation defined in (2.9) is the upper limit of integration changed from a constant to a variable. We first prove a more general fixed point theorem. In the following by $T^n$, we mean $T$ composed with itself $n$ times.

**Theorem 88** (Fixed Point)**.** *Let $X$ be a complete metric space and $T : X \to X$ a mapping such that $T^n$ is a contraction on $X$ for some positive integer $n$. Then $T$ has a unique fixed point $x \in X$.*

*Proof.* Since $T^n$ is a contraction on X, $T^n$ has a unique fixed point $x \in X$ so that $T^n x = x$. Noting that

$$T^n(Tx) = T^{n+1}x = T(T^n x) = Tx,$$

we see that $Tx$ is also a fixed point of $T^n$, but there can be only one, so $Tx = x$. Now, any fixed point $x^*$ of $T$ is also a fixed point of $T^n$ since

$$T^n x^* = T^{n-1}(Tx^*) = T^{n-1}x^* = \cdots = Tx^* = x^*.$$

It follows that $x$ is the only fixed point of $T$. □

**Example 80.** Let $T : C[0,1] \to C[0,1]$ be a linear operator given by

$$Tx(t) = \int_0^t x(s)ds. \tag{2.11}$$

Then

a) $T$ is not a contraction on $C[0,1]$.

b) $T$ has a unique fixed point $x \equiv 0$.

c) $T^n$ are contractions on $C[0,1]$ for all $n \geq 2$.

*Proof.* a) Notice that

$$||T1 - T0|| = ||T1|| = \sup_{t\in[0,1]} t = 1 = ||1 - 0||.$$

Therefore we have that $T$ is not a contraction on $C[0,1]$.

b) Notice that

$$Tx(0) = \int_0^0 x(s)ds = 0,$$

which implies that $Tx(0) = 0$ and so $T$ has a unique fixed point at the function $x \equiv 0$.

c) We will proceed here by induction. First we show the base case of $n = 2$:

$$\begin{aligned} T^2x(t) = \int_0^t Tx(s)\,ds &= \int_0^t \left(\int_0^s x(s)du\right) ds \\ &\leq ||x|| \int_0^t \int_0^s duds \\ &= \frac{t^2}{2}||x|| \leq \frac{1}{2}||x|| \implies ||T^2x|| \leq \frac{1}{2}||x||. \end{aligned}$$

By induction it is easy to show that

$$||T^n x|| \leq \frac{1}{n!}||x||.$$

Thus we have

$$||T^n x - T^n y|| = ||T^n(x-y)|| \leq ||T^n||\ ||x-y|| \leq \frac{1}{n!}||x-y||.$$

Therefore we have that $T^n$ are contractions on $C[0,1]$ for all $n \geq 2$. □

**Theorem 89** (Volterra Integral Equation Theorem)**.** *Suppose $v \in C[a,b]$ and $k \in C([a,b] \times [a,b])$. Then*

$$x(t) - \mu \int_a^t k(t,s)x(s)ds = v(t)$$

*has a unique solution $x \in C[a,b]$ for every $\mu$.*

*Proof.* We can write the Volterra integral equation as $x = Tx$ with $T : C[a,b] \to C[a,b]$ defined by

$$Tx(t) = v(t) + \mu \int_a^t k(t,s)x(s)ds.$$

Then we have

$$\begin{aligned} |Tx(t) - Ty(t)| &= |\mu| \left| \int_a^t k(t,s)[x(s) - y(s)]ds \right| \\ &\leq |\mu| \int_a^t |k(t,s)||x(s) - y(s)|ds \\ &\leq |\mu|M|(t-a)d(x,y) \quad \forall t \in [a,b], \end{aligned}$$

where we have used the facts that $|k(t,s)| \leq M$ and $|x(s) - y(s)| \leq d(x,y)$. We will now show by induction that

$$|T^n x(t) - T^n y(t)| \leq |\mu|^n M^n \frac{(t-a)^n}{n!} d(x,y) \quad \forall t \in [a,b]. \tag{2.12}$$

For $n = 1$ this was just proved, so assume that (2.12) holds for $n$. We will now show that (2.12) holds for $n+1$. So we have

$$\begin{aligned} |T^{n+1}x(t) - T^{n+1}y(t)| &= |\mu| \int_a^t k(t,s)[T^n x(s) - T^n y(s)]ds \\ &\leq |\mu| \int_a^t |k(t,s)||T^n x(s) - T^n y(s)|ds \\ &\leq |\mu| \int_a^t |k(t,s)||\mu|^n M^n \frac{(s-a)^n}{n!} d(x,y)ds \\ &\leq \frac{|\mu|^{n+1}M^{n+1}}{n!} d(x,y) \int_a^t (s-a)^n ds \\ &= \frac{|\mu|^{n+1}M^{n+1}}{n!} d(x,y) \left[ \frac{(s-a)^{n+1}}{n+1} \right]_a^t \\ &= \frac{|\mu|^{n+1}M^{n+1}}{(n+1)!} (t-a)^{n+1} d(x,y). \end{aligned}$$

So, we have from the induction that

$$\begin{aligned} d(T^n x, T^n y) &= \sup_{t\in[a,b]} |T^n x(t) - T^n y(t)| \\ &\leq \frac{|\mu|^n M^n (b-a)^n}{n!} d(x,y). \end{aligned}$$

So by letting

$$\alpha_n = \frac{|\mu|^n M^n (b-a)^n}{n!},$$

for any fixed $\mu$ and sufficient large $n$, we have that $\alpha_n < 1$ since $\lim_{n\to\infty} \alpha_n = 0$. Hence, $T^n$ is a contraction on $C[a,b]$ and the assertion follows from the Fredholm Integral Equation Theorem. □

**Remark 46.** If you would like to read more about fixed points, consult [6, 23, 32, 44]. Applications of fixed points to economics and game theory are studied in [13].

## Exercises

1.

a) Let $f : [1,2] \to [0,3]$ be a continuous function with $f(1) = 0$ and $f(2) = 3$. Show that $f$ has a fixed point.

b) Let $f : [a,b] \to [a,b]$ be a Lipschitz function with Lipschitz constant $0 < L < 1$. Show that $f$ has a unique fixed point.

2.

a) Give an example of a mapping $T$ of a complete metric space into itself with the property $d(Tx, Ty) < d(x,y)$ for all $x, y$ with $x \neq y$ which has no fixed point.

b) Show that $f : \mathbb{R} \to \mathbb{R}$ defined as $f(x) = \cos x$ is not a contraction but that the function $g(x) = \frac{99}{100} \cos x$ is a contraction.

c) Show that the *cosine* function is a contraction mapping on $[0,a]$ for any $1 \leq a < \frac{\pi}{2}$. Using the Banach Contraction Mapping Theorem, find the solution to the equation $x = \cos x$ correct to three decimal places.

3. Let $X = \{x \in \mathbb{Q} : x \geq 1\}$ and $f : X \to X$ be defined by $f(x) = \frac{x}{2} + \frac{1}{x}$.

a) Show that for all $x, y \in X$ one has $d(f(x), f(y)) \leq \frac{1}{2} d(x,y)$, where $d$ is the usual metric on real numbers.

d) Show that for this contraction $f$ there is no $x \in X$ for which $x = f(x)$. Does this contradict the Banach contraction mapping theorem?

4. Let $x_0 > 0$ be any positive number, and let $x_{n+1} = (1 + x_n)^{-1}$, prove that this sequence converges, and find its limit.

5. Show that the system of equations:

$$\begin{aligned} x_1 &= \frac{1}{4}x_1 - \frac{1}{4}x_2 + \frac{2}{15}x_3 + 3 \\ x_2 &= \frac{1}{4}x_1 + \frac{1}{5}x_2 + \frac{1}{2}x_3 - 1 \\ x_3 &= -\frac{1}{4}x_1 + \frac{1}{3}x_2 - \frac{1}{3}x_3 + 2 \end{aligned}$$

has a unique solution.

6. Consider $\mathbb{R}^n$ with the metric $d(x, z) = \sum_{i=1}^{n} |x_i - z_i|$. Let $T$ be a mapping from $\mathbb{R}^n$ to itself defined by the system of linear equations

$$y_i = \sum_{j=1}^{n} a_{ij}x_j + b_i,$$

where $i = 1, 2, ..., n$. Under what conditions $T$ is a contraction? What will be the condition if the above metric is replaced by the Euclidean metric $d(x, z) = \left[\sum_{i=1}^{n} |x_i - z_i|^2\right]^{\frac{1}{2}}$?

7. If $T$ is a mapping from a complete metric space $(M, d)$ into itself such that $T^m$ is a contraction mapping for some $m \in N$, then show that $T$ has a unique fixed point.

8.

a) Convert the following initial value problem (IVP) to an integral equation and set up an iteration scheme to solve it

$$\frac{dy}{dx} = 3xy \quad \text{where } y(0) = 1.$$

b) Given the initial value problem

$$f'(x) = 1 + x - f(x) \text{ for } \frac{-1}{2} \leq x \leq \frac{1}{2} \text{ where } f(0) = 1,$$

first show the mapping $T : C[\frac{-1}{2}, \frac{1}{2}] \to C[\frac{-1}{2}, \frac{1}{2}]$ defined by

$$Tf(x) = 1 + x + \frac{1}{2}x^2 - \int_0^x f(t)dt$$

is a contraction, then set up an iteration scheme to solve it.

9. If the partial derivative $\dfrac{\partial f}{\partial x}$ of $f$ exists and is continuous on a rectangle $R$ (see Picard's Theorem 86), show that $f$ satisfies a Lipschitz condition on $R$ with respect to second argument.

10. Let $f$ be a real valued, twice continuously differentiable function on $[a, b]$. Let $\widetilde{x}$ be a simple zero of $f$ in (a,b). Show that Newton's method defined by:

$$x_{n+1} = g(x_n) \text{ and } g(x_n) = x_n - \frac{f(x_n)}{f'(x_n)}$$

is a contraction in some neighborhood of $\widetilde{x}$, so that the iterative sequence converges to $\widetilde{x}$ for any $x_0$ sufficiently close to $\widetilde{x}$.

(Hint: note that $\widetilde{x}$ is a simple zero implies that $f'(x) \neq 0$ on some neighborhood $U$ of $\widetilde{x}$ where $U \subset [a, b]$ and $f''$ is bounded on $U$).

11. Show that the Fredholm integral equation $y(x) = 1 + \lambda \displaystyle\int_0^1 e^{x-t} y(t)\, dt$ has a solution for $|\lambda| < \dfrac{1}{e}$. Set up an iteration scheme to show that

$$\lim_{n\to\infty} y_n(t) = 1 - \frac{\lambda}{1-\lambda} e^x \left(\frac{1}{e} - 1\right).$$

12. Let $k(x, y)$ be a continuous function on the unit square $0 \leq x, y \leq 1$ satisfying $|k(x, y)| < 1$ for all $x$ and $y$. Show that there is a continuous function $f(x)$ on $[0, 1]$ such that one has

$$e^{x^2} = f(x) + \int_0^1 k(x, y) f(y) dy.$$

Can there be more than one such function $f$?

## 2.3 Modes of Convergence

### Pointwise and Uniform Convergence

Suppose that we are given a sequence of functions $f_1, f_2, \ldots$ all defined on a subset $D$ of $\mathbb{R}$. For each fixed point $x_0 \in D$ we form the numerical sequence $\{f_n(x_0)\}$. If $\{f_n(x_0)\}$ converges to $f(x_0)$ as sequence of real numbers for every $x_0 \in D$, we say $\{f_n\}$ *converges pointwise* to $f$ and write $f_n \to f$. We will soon see that desirable properties of the functions $f_n$, such as continuity and integrability, need not be preserved under pointwise convergence. However, under a stronger type of convergence known as *uniform convergence*, the situation is much more positive.

**Definition 74.** A sequence of functions $\{f_n\}$ is *uniformly convergent* to $f$ on the set $D$ if for each $\epsilon > 0$ there exists a number $N$ such that $|f_n(x) - f(x)| < \epsilon$ for all $x \in D$ whenever $n \geq N$. We use $f_n \rightrightarrows f$ to describe uniform convergence (Figure 2.12).

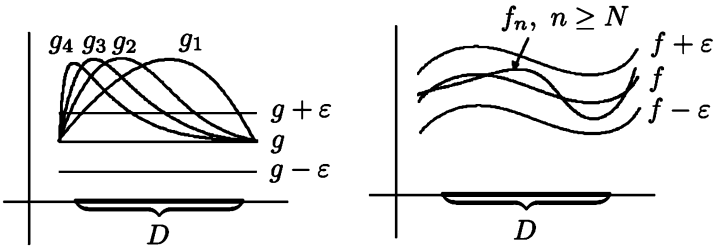



Figure 2.12: $g_n \to g$ pointwise but not uniformly, $f_n \rightrightarrows f$ uniformly

Note that uniform convergence does depend on the set $D$ being considered. Convergence could be uniform in one domain, but not on a larger domain. Clearly uniform convergence implies pointwise convergence; however, it imposes a stronger requirement on the number $N$ so that $N$ depends only on $\epsilon$ and not on the particular point $x_0 \in D$.

**Example 81.** The following are two examples of pointwise convergence.

a) The functions $f_n(x) = \dfrac{x}{n}$ converge pointwise to $f(x) \equiv 0$. Clearly these functions are just lines with decreasing slopes, but the convergence is not uniform on $\mathbb{R}$.

b) The functions $f_n(x) = x^n$ converge pointwise but not uniformly in the interval $[0, 1]$ to the limit $f(x) = 0$ for $x \in [0, 1)$ and $f(1) = 1$. Note that each of the functions $x^n$ is continuous; however, their limit $f$ is not (Figure 2.13). Furthermore, for $x > 1$, $\lim\limits_{n\to\infty} x^n$ does not exist and therefore for $x > 1$ $\{f_n\}$ does not converge pointwise. However, if we restrict the domain to $[0, 1]$, then

$$\lim_{n\to\infty} x^n = \begin{cases} 0 & \text{if } x \in [0, 1) \\ 1 & \text{if } x = 1 \end{cases}.$$

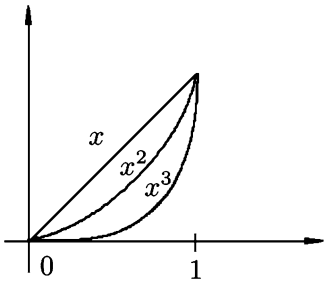



Figure 2.13: Limit function is not continuous

So we have pointwise convergence to $f$, however, $f$ is not continuous. The following theorem illustrates that continuity is preserved under uniform convergence.

**Theorem 90** (Interchange of limit operations)**.** *Suppose that $f_n$ for $n = 1, 2, \ldots,$ are defined and continuous on $D \subset \mathbb{R}$ such that $f_n \rightrightarrows f$ (uniformly) as $n \to \infty$. Then $f$ is continuous on $D$.*

*Proof.* Let $\epsilon > 0$ be given and let $x_0$ be any point in $D$. From

$$|f(x) - f(x_0)| = |f(x) - f_n(x) + f_n(x) - f_n(x_0) + f_n(x_0) - f(x_0)|,$$

the inequality

$$|f(x) - f(x_0)| \leq |f(x) - f_n(x)| + |f_n(x) - f_n(x_0)| + |f_n(x_0) - f(x_0)|$$

holds for all $x \in D$. Because $f_n \rightrightarrows f$, there exists a positive integer $N$ depending only on $\epsilon$ such that $|f(x) - f_n(x)| \leq \frac{\epsilon}{3}$ and $|f_n(x_0) - f(x_0)| \leq \frac{\epsilon}{3}$ for all $x \in D$ whenever $n \geq N$. Choose $n \geq N$ and use the fact that $f_n$ is continuous at $x_0$ to conclude $|f_n(x) - f_n(x_0)| \leq \dfrac{\epsilon}{3}$ holds. Therefore

$$|f(x) - f(x_0)| \leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon$$

holds for all $x \in D$ with $|x - x_0| < \delta$. Thus $f$ is continuous at $x_0$. □

Note that under conditions of the above theorem, when $\{x_m\}$ is any convergent sequence in $D$ whose limit $x$ is in $D$, then by the continuity of $f_n$'s we have $\lim_{m\to\infty} f_n(x_m) = f_n(x)$, therefore

$$\lim_{n\to\infty} (\lim_{m\to\infty} f_n(x_m)) = \lim_{m\to\infty} (\lim_{n\to\infty} f_n(x_m)).$$

Thus we see that limit operations can be interchanged.

**Definition 75.** Let $C[a,b]$ denote the space of all continuous real-valued functions defined on $[a,b]$. For $f \in C[a,b]$, define

$$||f|| = \sup\{|f(x)| : \ x \in [a,b]\}.$$

We say $f_n$ converges to $f$ in $C[a,b]$ if and only if $||f_n - f|| \to 0$.

**Theorem 91** (Uniform convergence in the sup norm)**.** *$\{f_n\}$ converges uniformly to $f$ on $[a,b]$ if and only if $\{f_n\}$ converges to $f$ in $C[a,b]$.*

*Proof.* Follows directly from definitions. □

## Test for Uniform Convergence

**Theorem 92** (Test for Uniform Convergence of Sequences)**.** *A sequence of functions with domain $D$ converges uniformly to a function $f$ if there is a real-valued sequence $\{a_n\}$ such that $a_n \to 0$ and*

$$|f_n(x) - f(x)| \leq |a_n|$$

*for all $x \in D$ and for all $n \in N$.*

*Proof.* We know that $a_n \to 0$. So, given $\epsilon > 0$ there exists $n \in \mathbb{N}$ such that if $n > N$, then $|a_n| < \epsilon$. This $N$ is independent of $x \in D$ and so

$$|f_n(x) - f(x)| \leq |a_n| < \epsilon$$

for all $x \in D$ whenever $n > N$. This implies that $\{f_n\}$ converges uniformly. □

**Theorem 93.** *Let $\{f_n\}$ be a sequence of functions defined on $A$. Then $f_n \rightrightarrows f$ uniformly on $B \subset A \Leftrightarrow$ the sequence of numbers*

$$d_n := \sup\{|f_n(x) - f(x)| : x \in B, n \in \mathbb{N}\}$$

*converges to 0.*

*Proof.* $\Rightarrow$ Suppose that $f_n \rightrightarrows f$ uniformly. Then, given $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that $|f_n(x) - f(x)| < \epsilon$ for all $n > N$ and for all $x \in B$. So, for $n > N$, $d_n < \epsilon$. Therefore, $d_n \to 0$.
$\Leftarrow$ Suppose $d_n \to 0$. Then we know that for sufficiently large $n$ and for all $x \in B$,

$$|f_n(x) - f(x)| < d_n < \epsilon.$$

So $f_n \rightrightarrows f$ uniformly. □

**Remark 47.** Note that Theorems 91, 92, and 93 are several interpretations of the fact that

$$f_n \rightrightarrows f \quad \text{uniformly} \quad \text{if and only if} \quad ||f_n - f|| \to 0.$$

For example for Theorem 92 we can just take $a_n = ||f_n - f||$.

**Example 82.** Let $f_n(x) = \dfrac{nx^2}{1+nx}$ and $A = [0,1]$. We claim that $\{f_n(x)\}$ converges uniformly.

Note that $\lim_{n\to\infty} f_n(x) = \lim_{n\to\infty} \dfrac{nx^2}{1+nx} = x$. So, we have

$$|f_n(x) - f(x)| = \left|\frac{nx^2}{1+nx} - x\right| = \left|\frac{x}{1+nx}\right|.$$

Since $d_n = \sup\left\{\left|\dfrac{x}{1+nx}\right| : x \in [0,1], n \in \mathbb{N}\right\}$, choose $d_n = \dfrac{1}{1+n}$ which goes to 0 as $n \to \infty$. Therefore, we have uniform convergence.

**Example 83.** When a sequence of integrable functions $\{f_n(x)\}$ converges pointwise to $f(x)$ on an interval $[a,b]$, the limit function $f$ need not to be integrable (Figure 2.14). For example, take $f_n(x) = ng_n(x)$ where $g_n(x)$ is the "triangle function" defined as:

$$g_n(x) = \begin{cases} nx & \text{if } x \in [0, 1/n] \\ 2 - nx & \text{if } x \in [1/n, 2/n] \\ 0 & \text{if } x \in [2/n, 1] \end{cases}.$$

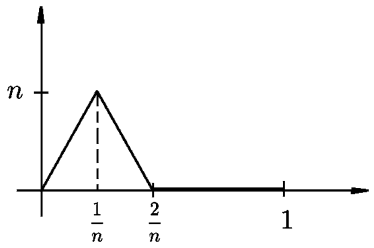



Figure 2.14: Graph of $f_n(x)$

The sequence $\{f_n\}$ converges pointwise (but not uniformly) to $f \equiv 0$, yet

$$\int_0^1 f_n(x)\,dx = 1, \text{ for every n, whereas } \int_0^1 f(x)\,dx = 0.$$

The next theorem shows that integrals behave nicely under uniform convergence.

**Theorem 94** (Interchange of limit and integral)**.** *If the functions $f_1, f_2, \ldots$ are integrable over an interval $[a,b]$ and if $f_n \rightrightarrows f$ (uniformly) on $[a,b]$, then $f$ is integrable and*

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} \int_a^b f_n(x)\,dx.$$

*Proof.* We again appeal to uniform convergence. Thus for all $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $n > N$, $|f_n(x) - f(x)| < \epsilon$. So we have

$$\left| \int_a^b f_n - \int_a^b f \right| = \left| \int_a^b f_n - f \right| \leq \int_a^b |f_n - f| < \epsilon(b-a).$$

It remains to show that $f$ is integrable. We need to show $U(f,P) - L(f,P) < \epsilon$ where $U(f,P)$ and $L(f,P)$ are upper and lower sums as used in the definition of the Riemann integral. However, for any partition $P$ of $[a,b]$,

$$\begin{aligned} U(f,P) - L(f,P) \quad \leq \quad & |U(f,P) - U(f_n,P)| + |U(f_n,P) - L(f_n,P)| \\ & + |L(f_n,P) - L(f,P)| \end{aligned}$$

holds true. Let $\epsilon > 0$, then by uniform convergence there is an $N$ such that the inequalities

$$|U(f,P) - U(f_n,P)| < \frac{\epsilon}{3} \quad \text{and} \quad |L(f_n,P) - L(f,P)| < \frac{\epsilon}{3}$$

hold true for all $n \geq N$ and for all partitions $P$ of $[a,b]$. On the other hand, integrability of $f_n$ allows for a partition P for which $|U(f_n,P) - L(f_n,P)| < \frac{\epsilon}{3}$. □

To see how uniform convergence behaves with respect to differentiation, consider the sequence of functions $f_n(x) = |x|^{1+\frac{1}{n}}$. Note that $f_n(x) \to f(x) = |x|$ and that $f_n \rightrightarrows f$ (uniformly) on $[-1,1]$. Even though each $f_n$ is differentiable, $f$ is not differentiable at zero. Thus while continuity is preserved, differentiability is not preserved under uniform convergence. Another common example is the sequence of functions $f_n(x) = \frac{1}{n}\sin(n^2 x)$. These functions are differentiable and converge uniformly to $f(x) \equiv 0$, but their derivatives $f_n'(x) = n\cos(n^2 x)$ do not even stay bounded if we take their limit as $n \to \infty$.

**Theorem 95** (Limit of a derivative is the derivative of the limit)**.** *Suppose $\{f_n\}$ is a sequence of continuous functions on an interval $[a,b]$ which have continuous derivatives on $(a,b)$. Assume that $f_n \to f$ for each $x \in [a,b]$ and the sequence of derivatives $f_n' \rightrightarrows g$ on $(a,b)$. Then the limit function $f$ is differentiable on $(a,b)$ and $g = f'$.*

*Proof.* Since $f_n' \rightrightarrows g$ and each $f_n'(x)$ is continuous we know that $g(x)$ is a continuous function on $(a,b)$. Furthermore, uniform continuity allows us to interchange the limit and integration, thus

$$\int_a^x g(t)dt = \lim_{n\to\infty} \int_a^x f_n'(t)dt = \lim_{n\to\infty} [f_n(x) - f_n(a)] = f(x) - f(a)$$

holds true for each $x \in (a,b)$. Thus, by the Fundamental Theorem of Calculus, $f$ is differentiable and $f'(x) = g(x)$ on $(a,b)$. □

## Dini's Theorem

Since pointwise convergence does not imply uniform convergence, one wonders under what additional conditions pointwise convergence gives uniform convergence. In the following we simply state a theorem called *Dini's Theorem* which provides answers to this question. For more details we refer the reader to [20], p. 39.

**Theorem 96** (Dini's Theorem)**.** *Let $A \subset \mathbb{R}$ be a compact set. Suppose $\{f_k\} : A \to \mathbb{R}$ is a sequence of continuous functions such that*

*a) $f_k(x) \geq 0$ for each $x \in A$.*

*b) $f_k(x) \leq f_\ell(x)$ whenever $k \geq \ell$.*

*c) $f_k \to 0$ pointwise for each $x \in A$.*

*Then $f_k \rightrightarrows 0$ (uniformly) on $A$.*

This theorem states that from pointwise convergence plus some other conditions we can conclude uniform convergence. Note also that Dini's theorem implies that if a sequence of continuous functions $f_n$ defined on $[a, b]$ converges monotonically to $f$ for each $x \in [a, b]$ and if $f$ is continuous, then the convergence is uniform. Suppose that a given sequence is monotonically increasing,

$$f_n(x) \leq f_{n+1}(x) \leq f(x).$$

Then by setting $g_n(x) = f(x) - f_n(x)$ we obtain a sequence of continuous functions $g_n$ that decreases pointwise to 0. Dini's Theorem then implies that $g_n \rightrightarrows 0$ as $n \to \infty$ in $[a, b]$. A similar argument works if the given sequence $\{f_n\}$ is monotonically decreasing.

**Definition 76** (Convergence of Series)**.** Consider an infinite series $\sum_{k=1}^{\infty} g_k(x)$ of functions defined on a set $A \subset \mathbb{R}$. We say this series converges to $g$ if the sequence of partial sums $\{s_n\}$, where $s_n = \sum_{k=1}^{n} g_k$, converges to $g$ pointwise, i.e., $s_n \to g$. If the convergence is uniform ( if $s_n \rightrightarrows g$) the infinite series converges uniformly.

The previous theorem has an equivalent expression for series.

**Corollary 15.** *Let $g_k : A \to \mathbb{N}$ be continuous and assume that $\sum_{k=1}^{\infty} g_k$ converges uniformly on a set $A$. Then its sum $g$ is continuous.*

This is true because continuity is preserved under uniform convergence and each $g_k$ is continuous.

**Corollary 16.** *If the sequence of functions $g_1, g_2, \ldots$ are integrable over an interval $[a,b]$ and the infinite series $\sum_{n=1}^{\infty} g_n \rightrightarrows g$ on $[a,b]$, then $g$ is integrable and*

$$\int_a^b g(x)\,dx = \sum_{n=1}^{\infty} \int_a^b g_n(x)\,dx.$$

This result is expressed roughly by indicating that a uniformly convergent series of functions may be integrated term by term, or that summing a series and then integrating the sum is equal to integrating each term and then summing up the integrals.

## Weierstrass M-test

The following is a very useful test by which a series of functions may sometimes be shown to be uniformly convergent.

**Theorem 97** (Weierstrass M-test)**.** *Let $\{f_k\}$ be a sequence of functions with domain $A$ and let there exist constants $M_k$ with*

$$|f_k(x)| \leq M_k$$

*for all $x \in A$ and $\sum_{k=1}^{\infty} M_k$ convergent. Then the series $\sum_{k=1}^{\infty} f_k$ converges uniformly and absolutely.*

*Proof.* Since we have $|f_k(x)| \leq M_k$ for all $x \in A$, we know that

$$\sum_{k=1}^{\infty} |f_k(x)| \leq \sum_{k=1}^{\infty} M_k.$$

So by comparison test this series converges absolutely and therefore it converges for each $x \in A$. Thus, the series $\sum f_k(x)$ converges pointwise to a function which we will call $s(x)$. Then we look at $s_n = f_1 + f_2 + \cdots + f_n$ and $t_n = M_1 + M_2 + \cdots + M_n$ where each $s_n$ is a function and each $t_n$ is a real number. Say that $s_n \to s$ and $t_n \to t$. Thus, for each $x$, if $n > m$ we have

$$|s_n(x) - s_m(x)| = \left| \sum_{k=m+1}^{n} f_k(x) \right| \leq \sum_{k=m+1}^{n} |f_k(x)| < \sum_{k=m+1}^{n} M_k = t_n - t_m.$$

Thus we have that the sequence

$$\{|s_n(x) - s_m(x)|\}_{n=1}^{\infty}$$

and $\{t_n - t_m\}$ are both convergent. If we take the limit as $n \to \infty$, then we know that $|s(x) - s_m(x)| \leq t - t_m$ for all $m \in \mathbb{N}$ and $x \in A$. When $m \to \infty$, $t - t_m \to 0$, and this gives the uniform convergence. $\square$

**Example 84.** a) The series $\sum_{n=1}^{\infty} \frac{(\sin(nx))^2}{n^2}$ converges uniformly. This can be deduced by considering the inequality:

$$\left|\frac{(\sin(nx))^2}{n^2}\right| \leq \frac{1}{n^2}.$$

Let $M_n = \frac{1}{n^2}$; since we know that $\sum \frac{1}{n^2}$ is a convergent $p$-series (with $p = 2$), by the Weierstrass M-test, the given series converges uniformly.

b) The series $\sum_{n=1}^{\infty} e^{-nx}$ converges uniformly on any closed subinterval of $(0, \infty)$.

Let $I = [a, \infty)$. If $x \in I$, then $e^{-nx} \leq e^{-na}$. We know that $\sum_{n=0}^{\infty} e^{-na}$ is a convergent geometric series. Therefore, by the Weierstrass M-test, the series converges uniformly.

**Remark 48.** A simple and useful test for uniform convergence called *Dirichlet's test* is as follows: Suppose that the $N$-th partial sum of the series $\sum f_n(x)$ is uniformly bounded with respect to both $x$ and $N$ on the interval $I$ and that $g_n(x)$ is a monotone decreasing sequence converging uniformly to zero. Then the series $\sum_{n=1}^{\infty} f_n(x)g_n(x)$ converges uniformly on $I$. (see, e.g., [26], p. 287, also Exercise 11 below).

The Cauchy criterion for convergence of a numerical sequence can be adapted to a criterion for uniform convergence.

**Definition 77** (Uniform Cauchy Criterion)**.** Suppose $\{f_n\}$ is a sequence of functions defined on $D \subset \mathbb{R}$. Then we say the sequence $f_n$ is a *uniform Cauchy sequence* on $D$ if for every $\epsilon > 0$, there exists $N \in \mathbb{N}$ such that $n, m > N$ implies that

$$|f_n(x) - f_m(x)| < \epsilon$$

for all $x \in D$.

It is not hard to observe that every uniform Cauchy sequence is uniformly convergent. To see this first apply the uniform Cauchy property to state

$$|f_n(x) - f_m(x)| < \epsilon \quad \text{for all} \quad x \in D$$

when $n, m \geq N$. Now, fix $n$ and take limit as $m \to \infty$. Since $f_m \to f$ as $m \to \infty$, we obtain that

$$|f_n(x) - f(x)| < \epsilon$$

for all $x \in D$ which implies that $f_n \rightrightarrows f$ uniformly. Note that the Weierstrass M-test can be obtained as an application of the uniform Cauchy criterion.

**Remark 49.** In some of the above proofs that involve integrals we use the following inequalities many times. Perhaps it is useful to list them:

a) If $f$ and $g$ are integrable functions on $[a, b]$, and $f(x) \leq g(x)$ for all $x \in [a, b]$, then

$$\int_a^b f(x)dx \leq \int_a^b g(x)dx.$$

b) If $f$ is integrable on $[a, b]$, so is the function $|f|$, and

$$\left| \int_a^b f(x)dx \right| \leq \int_a^b |f(x)|dx.$$

The above inequality (b) is the integral version of the well known triangle inequality $|a + b| \leq |a| + |b|$ for any numbers $a$ and $b$. If one assumes the integrability of $|f|$, then (b) follows from (a) and from the fact that

$$-|f(x)| \leq f(x) \leq |f(x)|.$$

**Remark 50.** There are many excellent texts in analysis and many contain the topics of uniform and pointwise convergence and their consequences. For example [1, 18, 26, 42, 52] are such texts.

## Weierstrass Function

An important application of uniform convergence, and in particular the use of the Weierstrass M-Test, is the following theorem. This theorem states a surprising fact that there exists a continuous function on $\mathbb{R}$ which is nowhere differentiable. The existence of a continuous function that has no tangent at any point is quite unintuitive. With this property, if you consider the graph of such a function and zoom in at an arbitrary point over and over again, a function will never become smooth or linear. The Weierstrass function is a jagged function known as a fractal curve.

**Theorem 98.** *There exists a real continuous function on the real line which is nowhere differentiable.*

*Proof.* For the proof of this theorem we refer the reader to page 154 of [42], where the construction of such a function is explained step by step. □

The original Weierstrass function is defined to be

$$f(x) = \sum_{n=0}^{\infty} a^n \cos(b^n \pi x),$$

where $0 < a < 1$, $b > 1$ is an odd positive integer and $ab > 1 + \frac{3}{2}\pi$. There are many other functions that also exhibit this property of being continuous

yet nowhere differentiable. Another one of these is $f(x) = \sum_{n=0}^{\infty} \frac{1}{2^n} \sin(2^n x)$. This function is also a Fourier series which has other interesting properties. Weierstrass original function not only gives a perspective about the relationship between continuity and differentiability, it also leads to the exploration of new set of functions.

## Exercises

1. Suppose that the sequence $\{f_n\}$ converges uniformly to $f$ on the set $D$ and that for each $n \in \mathbb{N}$, $f_n$ is bounded on $D$. Prove that $f$ is bounded on $D$.

2. Let $\{f_n\}$ be a sequence of functions with domain $A$. Show that if $\sum_{k=1}^{\infty} f_k$ is uniformly convergent in $A$, then $f_n \rightrightarrows 0$ on $A$.

3. Consider the sequence $\{f_n\}$ defined by $f_n(x) = \frac{nx}{1+nx}$, for $x \geq 0$.

   a) Find $f(x) = \lim_{n\to\infty} f_n(x)$.

   b) Show that for $a > 0$, $\{f_n\}$ converges uniformly to $f$ on $[a, \infty)$.

   c) Show that $\{f_n\}$ does not converge uniformly to $f$ on $[0, \infty)$.

4. Determine whether the sequence $\{f_n\}$ converges uniformly on $D$.

   a) $f_n(x) = \frac{1}{1+(nx-1)^2}$ $\qquad D = [0, 1]$.

   b) $f_n(x) = nx^n(1-x)$ $\qquad D = [0, 1]$.

   c) $f_n(x) = \arctan\left(\frac{2x}{x^2+n^3}\right)$ $\qquad D = \mathbb{R}$.

5. Suppose a sequence of functions $\{f_n\}$ is defined as

$$f_n(x) = 2x + \frac{x}{n} \qquad x \in [0, 1].$$

   a) Find the limit function $f = \lim_{n\to\infty} f_n$.

   b) Is $f$ continuous on [0,1]?

   c) Does $[\lim_{n\to\infty} f_n(x)]' = \lim_{n\to\infty} f_n'(x)$ for $x \in [0, 1]$?

   d) Does $\int_0^1 \lim_{n\to\infty} f_n(x)dx = \lim_{n\to\infty} \int_0^1 f_n(x)dx$?

6. Discuss the uniform convergence of the following series:

a) $\displaystyle\sum_{n=0}^{\infty} \frac{x^n}{n!}$ on $\mathbb{R}$

b) $\displaystyle\sum_{n=1}^{\infty} \frac{\sin(nx)}{\sqrt{n}}$ on $[0, 2\pi]$

c) $\displaystyle\sum_{n=1}^{\infty} \frac{\cos^2(nx)}{n^2}$ on $\mathbb{R}$

7. Let $f_n : [0,1] \to \mathbb{R}$ be a sequence of continuous functions such that

$$\int_0^1 \left(f_n(x) - f_m(x)\right)^2 \, dx \to 0 \quad \text{as } n, m \to \infty.$$

Let $k(x, y)$ be a continuous real-valued function on $[0,1] \times [0,1]$. Define

$$g_n(x) = \int_0^1 k(x,y) f_n(y) \, dy.$$

Prove that the sequence $\{g_n\}$ converges uniformly.
Hint: show that the sequence $\{g_n\}$ is a Cauchy sequence in the sup norm of $C[0,1]$.

8. Show that there exists a continuous function defined on $\mathbb{R}$ that is nowhere differentiable by proving the following:

a) Let $g(x) = |x|$ if $x \in [-1,1]$. Extend $g$ to be periodic. Sketch $g$ and the first few terms of the sum

$$f(x) = \sum_{n=1}^{\infty} \left(\frac{3}{4}\right)^n g(4^n x).$$

b) Use the Weierstrass M-test to show that $f$ is continuous.

c) Prove that $f$ is not differentiable at any point.

9. Can we differentiate the series

$$x = \sum_{k=1}^{\infty} \left(\frac{x^k}{k} - \frac{x^{k+1}}{k+1}\right), \quad 0 \le x \le 1$$

term by term?

10. Find an example of a sequence $\{f_k\}$ that converges uniformly to 0 on $[0,\infty)$, where each $\int_0^\infty f_k(x)dx$ exists, but $\int_0^\infty f_k(x)dx \to \infty$. Does this contradict Theorem 94?

11. Use Dirichlet's test to show that the trigonometric series $\sum_{n=1}^{\infty} \frac{\sin n\theta}{n}$ converges uniformly to $\frac{\pi-\theta}{2}$ on the interval $[\delta, 2\pi-\delta]$ for any $\delta > 0$.

12. Let $f$ be a continuous function on $\mathbb{R}$. Let

$$\frac{1}{n}\sum_{n=0}^{n-1} f\left(x+\frac{k}{n}\right).$$

Show that $\{f_n(x)$ converges uniformly to a limit on every finite interval $[a,b]$.

13. Let $f_n \in C[a,b]$ be a monotone increasing sequence $f_1(x) \le f_2(x) \le \cdots$ which converges pointwise to $f(x) \in C[a,b]$. Show that the convergence is uniform on $[a,b]$.

## 2.4 Approximation by Polynomials

The Weierstrass approximation theorem is one of the crowning results of classical analysis. It states that every continuous function on a closed and bounded interval can be approximated uniformly by algebraic polynomials, i.e., by functions of the form

$$P(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n \, eUnALT$$

where $n$ is a positive integer and the $a_k$'s are real numbers. This theorem is a prototype for a large collection of results in approximation theory and there are many different proofs of this theorem. We start by introducing the notion of *separability*.

**Definition 78.** A metric space $M$ is called *separable* if it contains a countable dense subset.

As we defined before, we say $A \subset M$ is *dense* if $\overline{A} = M$. Roughly speaking, a subset $A$ of a metric space $M$ is dense if the set $A$ union its accumulation points equals the metric space $M$. For example, the set $\mathbb{R}$ of real numbers is separable since the countable set $\mathbb{Q}$ is dense in $\mathbb{R}$. This is so, because as proved in the earlier chapter, every real number can always be given as the limit of a sequence of rational numbers, or, equivalently, because there always exists a rational number arbitrarily close to any given irrational number. There are other examples of separable spaces. These all depend on two facts: the set $\mathbb{Q}$ of

rational numbers is countable and the Cartesian product of any finite number of countable sets is again a countable set. Thus $\mathbb{R}^n$ and $\mathbb{C}^n$ are separable spaces. However, for the special form of the Weierstrass approximation theorem, we first need to ask the following questions:

## Dense Subsets of $C[a,b]$

a) Is $C[a,b]$ separable? Are there any "useful" dense subspaces of $C[a,b]$?

b) What are the compact subsets of $C[a,b]$?

As usual by $C[a,b]$ we mean the space of real-valued continuous functions defined on $[a,b]$. We supply $C[a,b]$ with the sup norm

$$||f||_\infty = \sup_{x\in[a,b]} |f(x)|.$$

We proved that convergence in $C[a,b]$ is the same as uniform convergence. Specifically,

$$f_n \to f \text{ in } C[a,b] \Leftrightarrow ||f_n - f||_\infty \to 0 \Leftrightarrow f_n \rightrightarrows f \text{ on } [a,b].$$

To answer these questions we first make the observation that $C[a,b]$ and $C[0,1]$ are in "some sense" identical. Thus we need only concern with a single choice of the interval $[a,b]$, and $[0,1]$ is often the most convenient.

**Lemma 9.** *There is a linear isometry from $C[0,1]$ onto $C[a,b]$ that maps polynomials to polynomials.*

*Proof.* Define $g : [a,b] \to [0,1]$ by

$$g(x) = \frac{x-a}{b-a} \quad \text{for} \quad a \le x \le b;$$

then $g$ is a homeomorphism. Note that a homeomorphism is a one-to-one, onto map for which both $g$ and $g^{-1}$ are continuous. In mathematics, we think of homeomorphic spaces as being essentially identical.

The map

$$T_g : C[0,1] \to C[a,b]$$
$$T_g(f) = f \circ g$$

defines a linear isometry from $C[0,1]$ onto $C[a,b]$. We observe the following about $T_g$:

1. Clearly $T_g$ is linear; it is one-to-one and onto. The last claim follows since $T_g$ has an inverse, namely

$$T_{g^{-1}}(h) = h \circ g^{-1} \quad \text{where} \quad g^{-1} : [0,1] \to [a,b]$$

with $g^{-1}(t) = a + t(b-a)$ for $0 \le x \le 1$.

2. $T_g$ is an isometry because $g$ is onto and

$$||T_g(f)||_\infty = \max_{x\in[a,b]} |f(g(x))| = \max_{t\in g([a,b])} |f(t)| = \max_{t\in[0,1]} |f(t)| = ||f||_\infty.$$

3. $T_g$ maps polynomials to polynomials.

If $p(t) = \sum_{k=0}^{n} a_k t^k$ is a polynomial, then

$$T_g(p(t)) = p(g(t)) = \sum_{k=0}^{n} a_k (g(t))^k = \sum_{k=0}^{n} a_k \left(\frac{t-a}{b-a}\right)^k$$

which is also a polynomial. □

**Proposition 29.** *$C[0,1]$ is separable.*

**Idea of the proof.** Let $f \in C[0,1]$ and let $\epsilon > 0$. We first approximate $f$ by a polygonal function shown in Figure 2.15.

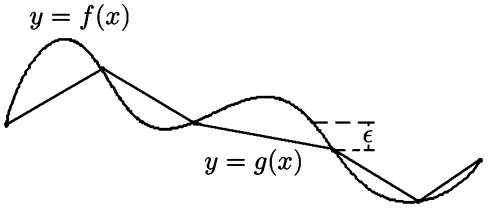



Figure 2.15: Approximation of $f$ by a polygonal function

Since $f \in C[0,1]$, $f$ is uniformly continuous, and we can find a sufficiently large $n$ so that

$$|f(x) - f(y)| < \epsilon \quad \text{whenever} \quad |x-y| < \frac{1}{n}.$$

Now define a polygonal function $g$ as

$$g\left(\frac{k}{n}\right) = f\left(\frac{k}{n}\right) \text{ for } k = 0, ..., n$$

and $g$ is linear on each interval $\left(\frac{k}{n}, \frac{k+1}{n}\right)$; moreover,

$$\|f-g\|_\infty = \max_{a\le t\le b} |(f-g)(t)| < \epsilon.$$

Next we change our approximating function to $h$. Let $h$ be another polygonal function such that

a) $h$ also has nodes at $\frac{k}{n}$ for $k = 0, 1, ..., n$.

b) $h\left(\frac{k}{n}\right)$ is rational.

c) $\left\|h\left(\frac{k}{n}\right) - g\left(\frac{k}{n}\right)\right\|_\infty < \epsilon$ for each $k$.

Then $\|h - g\|_\infty < \epsilon$ which will imply $\|f - h\|_\infty < 2\epsilon$. And the set of all polygonal functions taking only rational values at the nodes $\left(\frac{k}{n}\right)_{k=0}^{n}$ for some $n$ is countable. More precisely, for each $n$, let $Q_n$ be the set of all polygonal functions that have nodes at $x = \frac{k}{n}, k = 0, 1, ..., n$ and take only rational values at these points. To finish the proof, show that $Q_n$ is a countable set and $\bigcup Q_n$ is a countable dense set in $C[0, 1]$. □

## Weierstrass Approximation Theorem

If a function $f(x)$ is continuous on a closed and bounded interval $[a, b]$, then for each $\epsilon > 0$ there exists a polynomial $P(x)$ such that $|f(x) - P(x)| < \epsilon$ for all $x \in [a, b]$.

**Remark 51.** 1. Repeating the above theorem in the notation we have been using, given $f \in C[a, b]$ and $\epsilon > 0$, there is a polynomial $P(x)$ such that $\|f - P\|_\infty < \epsilon$. Hence there is a sequence of polynomials $(P_n)$ such that

$$P_n \rightrightarrows f \text{ on } [a, b].$$

2. The Weierstrass approximation theorem yields another proof that $C[a, b]$ is separable, meaning if $\mathcal{P}$ is the set of polynomials, then $\overline{\mathcal{P}} = C[a, b]$ where $\overline{\mathcal{P}}$ is the closure of $P$.

3. The theorem of Weierstrass is remarkable because continuous functions can have much more complicated behavior than polynomials. For example, recall Weierstrass' construction of continuous functions that are not differentiable at any point. If a function $f$ is sufficiently smooth, we can approximate it locally by partial sums of its Taylor series, but in general these polynomials do not provide the global approximation that the Weierstrass theorem provides.

## Bernstein Polynomials

**Definition 79.** Let $f \in C[0, 1]$. The polynomial $B_n(f)$ defined as

$$B_n(x, f) = (B_n(f))(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k} x^k (1 - x)^{n-k}, \quad 0 \le x \le 1$$

is called $n$th Bernstein polynomial of $f$.

Note that $B_n(f)$ is a polynomial of degree at most $n$ and

$$(B_n(f))(0) = f(0) \text{ and } (B_n(f))(1) = f(1).$$

Later we will show that $(B_n(f))(x)$ is in some sense an "average" of the numbers $f\left(\frac{k}{n}\right), k = 0, ..., n$.

**Example 85.** Let $f(x) = x^2$. To find $(B_1(f))(x)$ and $(B_2(f))(x)$ we set:

$$\begin{aligned}(B_1(f))(x) &= \sum_{k=0}^{1} f(k)\binom{1}{k}x^k(1-x)^{1-k} \\ &= f(0)\binom{1}{0}x^0(1-x)^1 + f(1)\binom{1}{1}x^1(1-x)^0 \\ &= x.\end{aligned}$$

$$\begin{aligned}(B_2(f))(x) &= \sum_{k=0}^{2} f\left(\frac{k}{2}\right)\binom{2}{k}x^k(1-x)^{2-k} \\ &= f(0)\binom{2}{0}x^0(1-x)^2 + f\left(\frac{1}{2}\right)\binom{2}{1}x^1(1-x)^1 \\ &\quad + f(1)\binom{2}{2}x^2(1-x)^0 \\ &= \frac{1}{4}\binom{2}{1}x(1-x) + \binom{2}{2}x^2 \\ &= 2x(1-x)\frac{1}{4} + x^2 = \frac{1}{2}x + \frac{1}{2}x^2.\end{aligned}$$

In fact one can find

$$(B_n(f))(x) = \frac{1}{n}x + \left(\frac{n-1}{n}\right)x^2$$

and

$$\lim_{n\to\infty} B_n(f) = x^2.$$

**Remark 52.** a) The binomial coefficient $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ and $0! = 1$ and the Binomial Theorem states that for $0 \le k \le n$,

$$(x+y)^n = \sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k}.$$

Thus if we substitute $y = 1 - x$ we obtain

$$1 = \sum_{k=0}^{n}\binom{n}{k}x^k(1-x)^{n-k}.$$

b) Bernstein's original proof is motivated by probability theory. Let $x$ be a fixed number in $[0,1]$. If a point is chosen randomly from $[0,1]$, it will lie in the interval $[0,x]$ with probability $x$ and in the interval $[x,1]$ with probability $1-x$. Another way of expressing is that probability of success is $x$, while probability of failure is $1-x$.

Thus the probability of exactly $k$ successes in $n$ independent trials is given by the binomial distribution

$$\binom{n}{k} x^k (1-x)^{n-k} := r_k(x).$$

Now consider a random variable $X_n$ that has a value $f\left(\frac{k}{n}\right)$. If $k$ of the chosen points are in $[0,x]$, the expectation (or mean) of $X_n$ is

$$E[X_n] = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) r_k(x) = (B_n(f))(x)$$

and it is possible to show $E[X_n] \to f(x)$ as $n \to \infty$ using some probabilistic arguments. However, we will concentrate on the following analytic proof. We refer the reader to [38] for a detailed study of Bernstein polynomials.

## Bernstein's Theorem

Let $f \in C[0,1]$ and let $\epsilon > 0$. Then there exists a polynomial $p(x)$ such that $||f-p|| < \epsilon$. In fact, the sequence of Bernstein polynomials

$$B_n(f) \rightrightarrows f \text{ ( converges uniformly to } f)$$

as $n \to \infty$.
The proof of Bernstein's Theorem follows from the following lemma.

**Lemma 10.** *Let $r_k(x)$ be as above, then*

$$\sum_{k=0}^{n} (k-nx)^2 r_k(x) = nx(1-x).$$

The proof of Lemma 10 will be given later. Observe that the sum above is a calculation of the binomial distribution.

**Proof of the Bernstein's Theorem.**

*Proof.* Observe that

$$\begin{aligned} |f(x) - (B_n(f))(x)| &= \left| f(x) - \sum_{k=0}^{n} f\left(\frac{k}{n}\right) r_k(x) \right| \\ &= \left| \sum_{k=0}^{n} \left( f(x) - f\left(\frac{k}{n}\right) \right) r_k(x) \right| \\ &\leq \sum_{k=0}^{n} \left| f(x) - f\left(\frac{k}{n}\right) \right| |r_k(x)|. \end{aligned}$$

To prove that $(B_n(f))(x)$ tends uniformly to $f(x)$ we appeal again to the uniform continuity of $f$ on $[0,1]$. Which means, for each $\epsilon > 0$, $\exists \delta > 0$ such that $|f(x) - f(y)| < \epsilon$ for all pairs of points $x, y \in [0,1]$ with $|x - y| < \delta$. Now choose $M$ such that $|f(x)| \leq M$ on $[0,1]$, and write

$$\begin{aligned} |f(x) - (B_n(f))(x)| &\leq \sum_{|k-nx|<\delta n} \left| f(x) - f\left(\frac{k}{n}\right) \right| |r_k(x)| \\ &\quad + \sum_{|k-nx|\geq\delta n} \left| f(x) - f\left(\frac{k}{n}\right) \right| |r_k(x)| \\ &\leq S_1 + S_2. \end{aligned}$$

In the above equation $S_1$ extends over all integers $k$ $(0 \leq k \leq n)$ for which $\left|x - \frac{k}{n}\right| < \delta$ and $S_2$ extends over all integers $k$ with $\left|x - \frac{k}{n}\right| \geq \delta$. By the uniform continuity of $f$ and the fact that $\sum_{k=0}^{n} r_k(x) = 1$ when $|k - nx| < \delta n$ we have $\left|f(x) - f\left(\frac{k}{n}\right)\right| < \epsilon$ and thus

$$S_1 \leq \sum_{k=0}^{n} \epsilon r_k(x) < \epsilon.$$

On the other hand, to get an upper bound for $S_2$ we will use the boundedness of $f$ on $[0,1]$ to conclude

$$\left| f(x) - f\left(\frac{k}{n}\right) \right| \leq |f(x)| + \left| f\left(\frac{k}{n}\right) \right| \leq 2M,$$

$$\begin{aligned} \Rightarrow S_2 &\leq 2M \sum_{|k-nx|\geq\delta n} r_k(x) = 2M \sum_{|k-nx|\geq\delta n} r_k(x) \frac{(k-nx)^2}{(k-nx)^2} \\ \Rightarrow S_2 &\leq \frac{2M}{\delta^2 n^2} \sum_{k=0}^{n} r_k(x) (k-nx)^2 \end{aligned}$$

and by the above Lemma 10

$$S_2 \leq \frac{2M}{\delta^2 n^2} nx\left(1-x\right).$$

Using the fact that maximum of the function $x\left(1-x\right)$ on $[0,1]$ is $\frac{1}{4}$ we obtain

$$S_2 \leq \frac{2M}{\delta^2 n}\frac{1}{4} < \frac{2M}{\delta^2 n} < \epsilon$$

for all $n \geq N > \frac{2M}{\delta^2 \epsilon}$. Note that $N$ depends only on $\epsilon$, so we have shown

$$\left|f\left(x\right) - \left(B_n\left(f\right)\right)\left(x\right)\right| \leq 2\epsilon$$

for all $x \in [0,1]$ whenever $n \geq N$. This shows

$$B_n\left(f\right) \rightrightarrows f \text{ on } [0,1]$$

which completes the proof of Bernstein's theorem. □

*Proof of Lemma 10.* By the binomial theorem we already observed that

$$1 = \left(x + \left(1-x\right)\right)^n = \sum_{k=0}^{n} \binom{n}{k} x^k \left(1-x\right)^{n-k} = \sum_{k=0}^{n} r_k\left(x\right).$$

Consider the binomial theorem again. Differentiating both sides with respect to $x$ yields:

$$\left(x+y\right)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

$$n\left(x+y\right)^{n-1} = \sum_{k=0}^{n} k\binom{n}{k} x^{k-1} y^{n-k}.$$

Then multiplying by $x$ gives:

$$nx\left(x+y\right)^{n-1} = \sum_{k=0}^{n} k\binom{n}{k} x^k y^{n-k}.$$

Setting $y = 1-x$ will give

$$nx = \sum_{k=0}^{n} k r_k\left(x\right).$$

Next we differentiate both sides of the binomial expansion twice and multiply by $x^2$. Thus we obtain

$$n\left(n-1\right)\left(x+y\right)^{n-2} x^2 = x^2 \sum_{k=0}^{n} k\left(k-1\right)\binom{n}{k} x^{k-2} y^{n-k}.$$

Substituting $y = 1 - x$ yields

$$\begin{aligned} n(n-1)x^2 &= \sum_{k=0}^{n} k(k-1) r_k(x) \\ &= \sum_{k=0}^{n} k^2 r_k(x) - \sum_{k=0}^{n} k r_k(x) \end{aligned}$$

$$\Rightarrow n(n-1)x^2 + nx = \sum_{k=0}^{n} k^2 r_k(x).$$

Now

$$\begin{aligned} \sum_{k=0}^{n} (k-nx)^2 r_k(x) &= \sum_{k=0}^{n} \left(k^2 - 2knx + n^2x^2\right) r_k(x) \\ &= \sum_{k=0}^{n} k^2 r_k(x) - 2nx \sum_{k=0}^{n} k r_k(x) + n^2x^2 \sum_{k=0}^{n} r_k(x) \end{aligned}$$

and we have expression for all of the above sums appearing on the right-hand side. Thus

$$\begin{aligned} \sum_{k=0}^{n} (k-nx)^2 r_k(x) &= nx + n(n-1)x^2 - 2nx(nx) + n^2x^2 \\ &= nx - nx^2 = nx(1-x) \quad \text{as desired.} \end{aligned}$$

□

**Remark 53.** Others proofs of the Weierstrass theorem are known. Landau's proof is indirect but closely follows the original ideas of Weierstrass; Lebesgue's proof is a very clever, simple proof. In the above we presented Bernstein's proof which was motivated by the theory of probability. For details of these proofs we refer the reader to [20], pp. 152–154. A nice introduction to approximation theory can be found in [15]; for a detailed discussion of the Weierstrass and approximation theory we refer the reader to [40] and a generalization of the Weierstrass approximation theorem can be found in M. H. Stone's paper [51].

## Exercises

1.

a) Show that $n$th Bernstein polynomial for $f(x) = e^x$ is

$$B_n(x) = [1 + (e^{\frac{1}{n}} - 1)x]^n.$$

b) Prove that $B_n(e^x)$ converges uniformly to $e^x$ on $[0, 1]$.

2. Show that the map $B_n$ is linear and monotone.

3. Prove that if $f$ is a continuous function on the closed and bounded interval $[a, b]$, then for any $\epsilon > 0$, there is a piecewise linear function $F$ which approximates $f$ uniformly within $\epsilon$ on the interval.

Hint: Divide the interval $[a, b]$ into $N$ equal subintervals at points

$$a = x_0 < x_1 < x_2 < \cdots < x_N = b.$$

Let $P_k$ be a point $(x_k, y_k)$ where $y_k = f(x_k)$ and define the function $F$ on each subinterval by

$$F(x) = \frac{(x_{k+1} - x)y_k + (x - x_k)y_{k+1}}{x_{k+1} - x_k}.$$

4.

a) Show that if $f \in C[a, b]$, and if $\displaystyle\int_a^b x^n f(x)dx = 0$ for each $n = 1, 2, ...,$ then $f = 0$.

b) For a $f \in C[a, b]$, the *moments* of $f$ are the numbers

$$\mu_n = \int_a^b x^n f(x)dx,$$

where $n = 0, 1, 2, ....$ Prove that two continuous functions defined on $[a, b]$ are identical if they have the same sequence of moments.

5. Apply the Weierstrass approximation theorem to show that if a function $f$ has a continuous derivative of order $n$ on an interval $[a, b]$, then for each $\epsilon > 0$ there is a polynomial $P$ such that

$$|f(x) - P(x)| < \epsilon, \ \ |f'(x) - P'| < \epsilon, \ldots, |f^n(x) - P^n(x)| < \epsilon$$

for all $x \in [a, b]$.

6. Let $f \in C[0, 1]$ have *Lipschitz constant 1*, then show that $||B_n f - f|| \leq \dfrac{1}{\sqrt{n}}$.

Recall that $f(x)$ defined on an interval $I$ is said to be *Lipschitz with constant* $L > 0$ provided that $|f(x) - f(y)| \leq L|x - y|$ for all $x, y \in I$.

7. Show that there exists a sequence of polynomials with integral coefficients converging to $f \in C[0, 1]$ uniformly if $f(0) = f(1) = 0$.

8. Deduce the approximation theorem by trigonometric polynomials from Weierstrass' approximation theorem by polynomials.

9. Suppose $f$ is a continuously differentiable function on $[0, 1]$ and $B_n f(x)$ is the Bernstein polynomial. Show that $B_n' f(x)$ converges uniformly to $f'(x)$ on the interval $[0, 1]$.

10. Given any function $f \in C[0, 1]$ and any number $\epsilon > 0$, show that there exists a polynomial $p$, all of whose coefficients are rational numbers, such that $||p - f|| < \epsilon$.

## 2.5 Functional Equations

Functional equations are equations that define their objects (functions) implicitly by their properties (equations) rather than by explicit formulas. Thus they are very much like algebraic equations except that the unknown quantities are functions rather than real numbers. There are some functional equations where one can guess functions that satisfy them. For example $f(x) - f(-x) = 0$ is satisfied by $f(x) = x^2$, $f(x) = \cos x$ or $f(x) = |-x|$, while the function $f(x) = 2x$ satisfies the functional equation $f(x+y) = f(x) + f(y)$, but the function $f(x) = x^2$ does not. This can be seen algebraically or by noting that the area $x^2 + y^2$ is a subset of the total area $(x+y)^2$ (Figure 2.16).

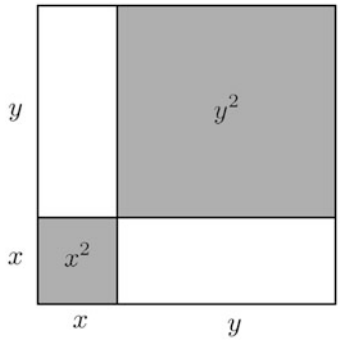



Figure 2.16: Area of $x^2 + y^2 \neq (x+y)^2$

Similarly, the functional equation $f(xy) = f(x) + f(y)$ is satisfied by all logarithmic functions while the functional equation $f(x+y) = f(x)f(y)$ is satisfied by all exponential functions. Given a functional equation, it is natural to ask the following questions: what function(s) are a solution to that equation? Or what values of the domain of the function satisfy the relationship given by the functional equation?

Sometimes more than one functional equation exists for a single function. For example there are three different functional equations for the gamma function $\Gamma(x)$, where the gamma function is defined as

$$\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt.$$

These are:

a) $\Gamma(x+1) = x\Gamma(x)$.

b) $\Gamma\left(\frac{x}{2}\right)\Gamma\left(\frac{x+1}{2}\right) = \frac{\sqrt{\pi}}{2^{x-1}}\Gamma(x)$ (Product formula).

c) $\Gamma(x)\Gamma(1-x) = \frac{\pi}{\sin(\pi x)}$ (Euler's functional equation).

**Theorem 99.** *(See [7], Theorem 6.3) The gamma function is the only continuous function that is positive for positive $x$ and that satisfies the functional equations:*

*a)* $\Gamma(x+1) = x\Gamma(x)$.

*b)* $$\Gamma\left(\frac{x}{p}\right)\Gamma\left(\frac{x+1}{p}\right)\Gamma\left(\frac{x+p-1}{p}\right) = \frac{(2\pi)^{(p-1)/2}}{p^{x-1/2}}\Gamma(x),$$

*for all values of $p$. Note that for $p = 2$ the equation above is the product formula.*

The study of functional equations which have an easy appearance can get quite complicated; at the same time in some cases "clever" tricks are employed to solve the equations. Perhaps that is why we find questions involving functional equations among Math Olympiads problems. There are books dedicated to finding solutions of functional equations [45]. There are two directions concerning inquiries for any given functional equation, say $\mathcal{F}(f) = 0$. One direction deals with *regularity*, which is proving under some assumption on $f$ that if $\mathcal{F}(f) = 0$, then $f$ is regular ($f$ is continuous, differentiable, analytic, etc.). The other direction concerns *stability* which is to prove that if $f$ is not a solution of $\mathcal{F}(f) = 0$, then there exists a unique function $g$ which is "near" $f$ and $\mathcal{F}(g) = 0$, meaning $f$ almost solves the given functional equation.

Finding a solution to functional equations is one of the oldest topics of mathematical analysis. D'Alembert, Euler, Gauss, Cauchy, Abel, Weierstrass, Darboux, and Hilbert are among the renowned mathematicians who were concerned with methods of solutions for functional equations. Some trace the earliest works on functional equations to N. Oresme in the 14th century who was the first mathematician to use functional equations in several variables. He considered the following functional equation:

$$\frac{f(x_1) - f(x_2)}{f(x_2) - f(x_3)} = \frac{x_1 - x_2}{x_2 - x_3}$$

for all $x_1, x_2, x_3$ with $x_1 > x_2 > x_3$. If one picks the point $x_2$ halfway between $x_1$ and $x_3$ one gets *Jensen's functional equation*:

$$f\left(\frac{x_1 + x_3}{2}\right) = \frac{f(x_1) + f(x_3)}{2}.$$

(For details see [2]).

## Cauchy's Equation

It is generally accepted that Cauchy (1789–1857) brought rigor to analysis. It is in his book *Cours d'analyse de l'Ecole Royale Polytechnique* that we first find systematic treatment of some basic functional equations. The equations

$$\begin{aligned} f(x+y) &= f(x) + f(y), \\ f(xy) &= f(x) + f(y), \\ f(x+y) &= f(x)f(y), \text{ and} \\ f(xy) &= f(x)f(y), \end{aligned}$$

were studied by Cauchy and the functional equation $f(x+y) = f(x) + f(y)$ is called *Cauchy's functional equation*. If a function $f : \mathbb{R}^n \to \mathbb{R}$ satisfies Cauchy's functional equation for every $x, y \in \mathbb{R}$, it is called *additive*. Using induction one can easily see that if a function $f$ is additive, then for all $x_1, x_2, \ldots, x_n \in \mathbb{R}$ for $f : \mathbb{R}^n \to \mathbb{R}$ one also has:

$$f(x_1 + x_2 + \cdots + x_n) = f(x_1) + f(x_2) + \cdots + f(x_n).$$

**Theorem 100.** *If $f : \mathbb{R}^n \to \mathbb{R}$ satisfies Cauchy's functional equation, then*

$$f(rx) = rf(x)$$

*for all $x \in \mathbb{R}^n$ and $r \in \mathbb{Q}$.*

*Proof.* Putting $x = x_1 = x_2 = \cdots = x_n$ into $f(x_1 + x_2 + \cdots + x_n) = f(x_1) + f(x_2) + \cdots + f(x_n)$, we get

$$f(nx) = nf(x).$$

Thus, if $x = \left(\frac{m}{n}\right) t$ then $f(nx) = f(mt)$; that is $f\left(\frac{m}{n}t\right) = \left(\frac{m}{n}\right) f(t)$. Letting $t = 1$ and $f(1) = c$, we obtain

$$f(x) = cx \tag{2.13}$$

for every positive rational $x$.
For $x = 0$, $f(0) = f(x + (-x)) = f(x) + f(-x) = 0$; thus $f$ is an odd function. From the equation $f\left(\frac{m}{n}\right) t = \frac{m}{n} f(t)$ we obtain

$$f(rt) = rf(t) \quad \text{for all real } t, \text{and all rational } r;$$

if $t = 1$, $f(r) = cr$ for all rational $r$. If $f(x)$ is assumed to be continuous everywhere, then it follows by taking limits of both sides of (2.13) that

$$f(x) = cx$$

for all real $x$. □

The above theorem basically says that every additive function $f : \mathbb{R}^n \to \mathbb{R}$ is a homomorphism from $\mathbb{R}^n$ to $\mathbb{R}$ both considered as linear spaces over $\mathbb{Q}$.

Note that every linear function of the form $f(x) = cx$ is additive, thus it is natural to ask if every additive function is a linear function of the form $f(x) = cx$. The answer to this question is negative. However, by imposing a continuity condition on $f$, we can show that $f(x) = cx$.

**Theorem 101.** *(Cauchy, 1821) If $f : \mathbb{R} \to \mathbb{R}$ is continuous everywhere and additive, then $f(x) = cx$.*

However, in 1875, G. Darboux observed that continuity only at a single point $x_0$ is sufficient, since in this case it follows that $f(x)$ is continuous everywhere. To see this, observe that continuity of $f$ at $x_0$ is expressed as $\lim_{s\to x_0} f(s) = f(x_0)$; then for any $x$ we want to show $\lim_{z\to x} f(z) = f(x)$. However,

$$\lim_{z\to x} f(z) = \lim_{z-x+x_0\to x_0} f[(z-x+x_0)+(x-x_0)] = \lim_{s\to x_0} f[s+(x-x_0)].$$

Using the additivity of $f$ then yields:

$$\lim_{z\to x} f(z) = \lim_{s\to x_0} f(s)+f(x-x_0) = f(x_0)+f(x-x_0) = f(x_0+x-x_0) = f(x).$$

In the following we prove a different version of Cauchy's theorem by assuming only boundedness of $f$ on an interval.

**Theorem 102.** *Let $f$ be an additive function and suppose there is an interval $[a, b]$ on which $f$ is bounded either from above or below. Then $f(x) = cx$.*

*Proof.* Assume $f$ is both additive and bounded above on $[a, b]$. The proof is similar if $f$ is bounded below on $[a, b]$. We want to show $f(x) = cx$ for every $x$. To accomplish this we set up a new function

$$h(x) = f(x) - cx$$

and observe its properties. First of all if we set $c = f(1)$, then $h(1) = 0$ and $h$ is additive. Note also that $h(r) = 0$ for every rational number for every rational number $r \in \mathbb{Q}$. If $n$ is a positive integer, then

$$0 = h(1) = h\left(\frac{1}{n}n\right) = nh\left(\frac{1}{n}\right),$$

thus Since

$$h\left(\frac{m}{n}\right) = mh\left(\frac{1}{n}\right) = 0$$

for every $m = 1, 2, \ldots$, $h$ is vanishing at the positive rationals. But since $h$ is an odd rationals. But since $h$ is an odd function we obtain that $h(r) = 0$ for all $r \in \mathbb{Q}$. The function $h$ is also bounded:

$$h(x) = f(x) - cx \le M + |c| \max(|a|, |b|) = K.$$

We can even claim that $h(x) \le K$ everywhere (just choose $r \in [a-x, b-x]$ and use that $h$ is $r$-periodic to conclude $h(x) = h(x+r) \le K$). All we need to show now is that $h(x) \le 0$ everywhere. But $h(nx) = nh(x) \le K$ for every $x$ and every $n = 1, 2, \ldots$ is true only when $h(x) \le 0$. This concludes the proof, since $h(x)$ is also an odd function; $h(-x) = -h(x) \le 0$ for every $x$ implies $h(x) \ge 0$ everywhere and hence $h(x) \equiv 0$. □

Note that there were improvements of the above theorem, and one such result of Kormes in 1926 states that if we assume $f$ is both additive and measurable, then $f(x) = cx$.

**Example 86.** Suppose a given function $f : \mathbb{R} \to \mathbb{R}$ is a function which satisfies both

a) $f(x + y) = f(x) + f(y)$,

b) $f(xy) = f(x)f(y)$

for all $x, y \in \mathbb{R}$. Then, either $f(x) \equiv 0$ or $f(x) \equiv x$.
Note that $f$ is bounded from below in $[0, 1]$, since for $x \geq 0$, we have

$$f(x) = f(\sqrt{x}\sqrt{x}) = f^2(\sqrt{x}) \geq 0.$$

Since $f$ is additive, using the above theorem, we can conclude that $f(x) = cx$ for some constant $c$. Then $f(1) = f^2(1)$ gives $c = c^2$, thus either $c = 0$ or $c = 1$.

Next we show that not every additive function is of the form $f(x) = cx$. We need the following concept called a *Hamel basis*.

**Definition 80.** A set $H \subset \mathbb{R}$ is called a *Hamel basis* if for every real number $x$, there exists a natural number $n \in \mathbb{N}$, $r_i \in \mathbb{Q}$ and $h_i \in H$ where $i = 1, 2, \ldots n$, such that $x$ can be written uniquely as:

$$x = r_1h_1 + r_2h_2 + \cdots + r_nh_n.$$

Observe that this definition is just a restatement of the concept of an algebraic basis for a vector space. The proof of the following proposition depends on the existence of a Hamel basis.

**Proposition 30.** *Not every additive function is necessarily of the form $f(x) = cx$.*

*Proof.* We claim that if $a$ and $b$ are nonzero real numbers such that $\dfrac{a}{b} \notin \mathbb{Q}$, then we can find an additive function $f$ with the properties that $f(a) \neq 0$ and $f(b) = 0$. Thus, such a function cannot have the form $f(x) = cx$. Since $\dfrac{a}{b}$ is irrational, the set $\{a, b\}$ is a linearly independent set over the rationals and it is contained in any Hamel basis $H$. We define the function $f$ as:

$$f(x) = \begin{cases} 0 & \text{if } a \text{ does not occur among } h_1, h_2, \ldots, h_n \\ \text{coefficient of } a & \text{in the unique representation of } x \text{ given above.} \end{cases}$$

Then the function $f$ is well defined and moreover for $x, y \in \mathbb{R}$, if

$$y = t_1h_1 + t_2h_2 + \cdots + t_nh_n, \quad \text{then}$$

$$x + y = (r_1 + t_1)h_1 + \cdots + (r_n + t_n)h_n$$

is the unique representation of $x + y$. We now consider the following cases:

a) If $a = h_i$, then $f(x+y) = r_i + t_i = f(x) + f(y)$.

b) If $a$ does not occur among $h_i$'s for $1 \leq i \leq n$, then

$$f(x) = 0 = f(y) = f(x+y).$$

In either case $f$ satisfies Cauchy's equation $f(x+y) = f(x) + f(y)$. □

**Remark 54.** In 1906 G. Hamel was the first mathematician to prove the existence of discontinuous additive functions. Discontinuous additive functions are sometimes called *Hamel functions*. Since the very first theorem we proved in this section tells us that every additive function $f : \mathbb{R}^n \to \mathbb{R}$ is a homomorphism from $\mathbb{R}^n$ to $\mathbb{R}$ where both $\mathbb{R}^n$ and $\mathbb{R}$ are considered as linear spaces over $\mathbb{Q}$, we have the following theorem as an immediate consequence.

**Theorem 103.** *Let $H$ be an arbitrary Hamel basis for $\mathbb{R}^n$. Then for every function $g : H \to \mathbb{R}$ there exists a unique additive function $f : \mathbb{R}^n \to \mathbb{R}$ such that $f|_H = g$ ($f$ restricted to $H$ is equal to $g$).*

As a corollary to this theorem, we can prove:

**Corollary 17.** *Let $H$ be an arbitrary Hamel basis for $\mathbb{R}$, and let $g : H \to \mathbb{R}$ be an arbitrary function. Let $f : \mathbb{R} \to \mathbb{R}$ be the unique extension of $g$. Then $f$ is continuous if and only if $\dfrac{g(x)}{x} = $ constant for $x \in H$.*

*Proof.* Continuity of $f$ implies that $f(x) = cx$ for $x \in \mathbb{R}$. In particular, for $x \in H$, we have $g(x) = f(x) = cx$. Conversely, if $g(x) = cx$ for $x \in H$, then the function $cx$ is an additive extension of $g$, and by the uniqueness of such extensions $f(x) = cx$. Thus $f$ is continuous. □

Now we see that it is enough to take any function $g : H \to \mathbb{R}$ such that

$$\frac{g(x)}{x} \neq \text{constant}$$

on $H$, and the additive extension of $g$ defines a discontinuous additive function.

**Remark 55.** If $f$ is additive and $f(x) \neq cx$ for any constant $c$, then we can find $x_1, x_2$ such that

$$\frac{f(x_1)}{x_1} \neq \frac{f(x_2)}{x_2},$$

which means that the determinant of the following matrix

$$\begin{pmatrix} x_1 & f(x_1) \\ x_2 & f(x_2) \end{pmatrix}$$

is equal to $x_1 f(x_2) - x_2 f(x_1) \neq 0$. Thus the vectors $\{(x_1, f(x_1)),\ (x_2, f(x_2))\}$ form a linearly independent set over $\mathbb{Q}$. However, if we look at a linear combination of these vectors, namely

$$r_1(x_1, f(x_1)) + r_2(x_2, f(x_2)) = (r_1 x_1 + r_2 x_2, f(r_1 x_1 + r_2 x_2))$$

for $r_1, r_2 \in \mathbb{Q}$, we get a point belonging to the graph $G(f)$ of $f$. This gives an idea why the following theorem is valid:

**Theorem 104** (Hamel, 1906 and San Juan, 1946)**.** *If $f : \mathbb{R} \to \mathbb{R}$ is additive and discontinuous, then*

$$\overline{G(f)} = \mathbb{R}^2,$$

*i.e., the closure of the graph of $f$ in $\mathbb{R}^2$ is equal to $\mathbb{R}^2$.*

## Jensen's Equation

Jensen's equation for convex functions is the functional equation

$$f\left(\frac{x+y}{2}\right) = \frac{f(x) + f(y)}{2}.$$

If we assume Jensen's equation is valid for all real $x, y$, then it is reduced to Cauchy's functional equation. If we set $y = 0$ in Jensen's equation, we obtain

$$f\left(\frac{x}{2}\right) = \frac{f(x) + f(0)}{2} = \frac{f(x) + c}{2},$$

where $f(0) = c$. Thus using Jensen's equation again we obtain

$$\frac{f(x) + f(y)}{2} = f\left(\frac{x+y}{2}\right) = \frac{f(x+y) + c}{2},$$

that is

$$f(x+y) = f(x) + f(y) - c.$$

Setting $g(x) = f(x) - c$, we see that the function $g$ satisfies Cauchy's equation, $g(x+y) = g(x) + g(y)$. Thus if $f$ is continuous at a point then $g(x) = cx$. We just proved the following theorem.

**Theorem 105.** *The function $f(x) = ax + c$ is the most general solution of the functional equation*

$$f\left(\frac{x+y}{2}\right) = \frac{f(x) + f(y)}{2}$$

*that holds for all real $x, y$ and that is continuous at one point.*

Note that above argument can be repeated to conclude the following corollary.

**Corollary 18.** *The most general solution of Jensen's functional equation is of the form*

$$f(x) = h(x) + a,$$

*where $a$ is an arbitrary constant and $h$ is an additive function satisfying $h(x + y) = h(x) + h(y)$.*

If we assume that $f$ is continuous only on a closed interval without paying attention to the behavior of the function outside this interval, we can still get information about the solution of Jensen's functional equation.

**Proposition 31.** *Suppose $f$ is a continuous function on $[0, 1]$. Then*

$$f(x) = cx + a$$

*is a solution of Jensen's equation.*

*Proof.* Setting $f(0) = a$, $f(1) = b$ and using

$$f\left(\frac{x+y}{2}\right) = \frac{f(x) + f(y)}{2},$$

we obtain

$$f\left(\frac{1}{2}\right) = f\left(\frac{0+1}{2}\right) = \frac{f(0) + f(1)}{2} = a + \frac{1}{2}(b - a).$$

Next, we look at $f\left(\frac{1}{4}\right)$ and $f\left(\frac{3}{4}\right)$:

$$f\left(\frac{1}{4}\right) = \frac{f(0) + f(1/2)}{2} = a + \frac{1}{4}(b - a),$$

$$f\left(\frac{3}{4}\right) = \frac{f(1/2) + f(1)}{2} = a + \frac{3}{4}(b - a).$$

Thus we claim that

$$f(x) = a + (b - a)x$$

holds for all dyadic $x$. For dyadic fractions with denominator $2^n$, we have:

$$f\left(\frac{2m}{2^{n+1}}\right) = f\left(\frac{m}{2^n}\right) = c + \frac{m}{2^n} + a = c + \frac{2m}{2^{n+1}} + a$$

and

$$f\left(\frac{2m+1}{2^{n+1}}\right) = \frac{f(m/2^n) + f(m + 1/2^n)}{2} = c + \frac{2m+1}{2^{n+1}} + a,$$

which proves the claim. Since $f$ is continuous we have already shown that $f(x) = cx + a$ for all $x \in [0, 1]$, where $c = b - a$. □

**Remark 56.** Let $A \subset \mathbb{R}^n$ be an open set. A function $f : A \to \mathbb{R}$ is called *convex* if and only if it satisfies Jensen's functional inequality:

$$f\left(\frac{x+y}{2}\right) \leq \frac{f(x)+f(y)}{2}$$

for all $x \in A$. For example $f(x) = x^2$, $f(x) = e^x$ and $f(x) = |x|$ are all convex functions on any open interval of $\mathbb{R}$. It turns out that the notion of convex functions is just as important as positive or increasing functions. That is probably why the treatment of convex functions can be found in many calculus books. In the following we just mention the relationship of convex functions and additive functions.

**Theorem 106.** *Every additive function is convex.*

*Proof.* For any additive function $f : \mathbb{R}^n \to \mathbb{R}$, we have

$$f\left(\frac{x+y}{2}\right) = \frac{1}{2}f(x+y) = \frac{f(x)+f(y)}{2}. \qquad \square$$

Note that every additive function is also concave, i.e., it satisfies

$$f\left(\frac{x+y}{2}\right) \geq \frac{f(x)+f(y)}{2}.$$

Thus additive functions are both convex and concave. The converse is not true, since functions of the form $f(x) = h(x) + b$ with $h$ an additive function and $b$ any constant are concave and convex, but not additive if $b \neq 0$. If a function $f : \mathbb{R}^n \to \mathbb{R}$ is at the same time convex and concave, then it differs from an additive function by a constant (for details see [36], Chapter 13).

## Exercises

1. Find all functions $f$ satisfying

$$f(x+y)^2 = f(x)^2 + f(y)^2$$

for all $x, y \in \mathbb{R}$.

2. Let $f : \mathbb{R} \to \mathbb{R}$ be a function satisfying Jensen's functional equation

$$f\left(\frac{x+y}{2}\right) = \frac{f(x)+f(y)}{2}$$

for every $x, y \in \mathbb{R}$. Show that there is a constant $c$ such that $f(x) - c$ is an additive function.

3. If the function $f : \mathbb{R} \to \mathbb{R}$ satisfies

$$x + f(x) = f(f(x)) \quad \text{for every} \quad x \in \mathbb{R},$$

find all solutions of the equation $f(f(x))$.

4. Prove that if $u = (a, b)$, $v = (c, d)$ form a linearly independent set in $\mathbb{R}^2$, then $W = \text{span}_{\mathbb{Q}}\{u, v\} = \{ru + sv : \ r, s \in \mathbb{Q}\}$ forms a dense subset of $\mathbb{R}^2$.

5. Define and solve Jensen's functional equation for continuous functions defined on any interval $[a, b]$, where $0 < a < b$.

6. Let $H$ be a Hamel basis and let $f : H \to \mathbb{R}$ be an arbitrary function. Prove that $f$ can be extended uniquely to $\mathbb{R}$ as an additive function.

7. The Gamma Function is defined by

$$\Gamma(p + 1) = \int_0^\infty e^{-x} x^p dx.$$

a) Show that for $p > 0$, $\Gamma(p + 1) = p\Gamma(p)$.

b) Show that $\Gamma(1) = 1$.

c) If $p$ is a positive integer $n$, show that $\Gamma(n + 1) = n!$.

8. Suppose that $f(x)$ is convex on an interval $I$ in the sense of Jensen. Show that

$$f\left(\frac{x_1 + x_2 + \cdots + x_n}{n}\right) \leq \frac{f(x_1) + f(x_2) + \cdots + f(x_n)}{n}$$

for arbitrary $n$-points $x_1, x_2, \ldots, x_n$ in $I$.

9. Show that $f(x)$ is convex on an interval $I$ if and only if $e^{\lambda f(x)}$ is convex on $I$ for any positive $\lambda$.

10. Show that $f \in C[a, b]$ is convex if and only if $\dfrac{1}{t - s} \displaystyle\int_s^t f(x)dx \leq \frac{f(x) + f(y)}{2}$ for any $s \neq t$ in $[a, b]$.

## 2.6 Fourier Series

In this section we focus on an introduction to Fourier Series. It is well known that in analysis, power series expansions of functions play an important role. Recall that to determine a Taylor series expansion of a function, we need to find the coefficients and the coefficients are determined by differentiation. But what if the function is not differentiable or what if it has points of discontinuity? Can we still have another expansion where the coefficients in this representation are determined by integration? (Since we have not developed measure theory or Lebesgue integration, what we mean by integration here is Riemann integrable functions). The answer to this question lies in the theory of Fourier Analysis. Since the time of Fourier (1768–1830), an extensive mathematical theory of Fourier series has emerged. Today, besides being a vibrant area of research of pure mathematics, it has broad applications to fields such as mathematical physics, medical imaging, data storage and retrieval, and coding theory. A Fourier series is an expansion of a periodic function $f(x)$ in terms of an infinite sum of sines and cosines. Fourier series make use of the orthogonality relationships of the sine and cosine functions, as we shall explain shortly. The computation and study of Fourier series is known as harmonic analysis and is extremely useful as a way to break up an arbitrary periodic function into a set of simple terms that can be plugged in, solved individually, and then recombined to obtain the solution to the original problem.

### Historical Background

The theory of Fourier series has its roots in the problems of mathematical physics. The story begins in 1746 with a famous *vibrating string problem*. Consider an elastic string of length $L$ which has each end fastened to one of the end points of the interval $[0, L]$ on the $x$-axis and set into motion (just like when you pluck a guitar string). The question is to determine the position $y = F(x, t)$ of the string at a time $t$ provided its initial position $F(x, 0) = f(x)$ and initial velocity $F_t(x, 0) = 0$. The function $F(x, t)$ is the solution of the so-called *d'Alembert's wave equation*:

$$F_{tt} = a^2 F_{xx},$$

where the initial data for this problem is

$$F(x, 0) = f(x), \quad F_t(x, 0) = 0, \quad f(0) = 0 = f(L),$$

and $a$ is a positive constant determined by the physical properties of the string. It is not hard to find particular solutions to the wave equation, for example each of the functions

$$F(x, t) = \sin\left(\frac{k\pi x}{L}\right) \cos\left(\frac{ak\pi t}{L}\right) \quad k = 1, 2, \ldots$$

is a solution. Note that the initial position is $F(x,0) = \sin\left(\frac{k\pi x}{L}\right)$. Moreover, if one assumes "superposition" of solutions (or equivalently if one assumes that we can differentiate term by term) we can suggest that any sum of the type

$$F(x,t) = \sum_{k=1}^{\infty} a_k \sin(\frac{k\pi x}{L}) \cos(\frac{ak\pi t}{L}) \tag{2.14}$$

is also a solution. If we accept that the equation (2.14) is the most general solution then we must also accept that the initial position $f$ must be

$$f(x) = \sum_{k=1}^{\infty} a_k \sin(\frac{k\pi x}{L}). \tag{2.15}$$

However, the sum in the equation (2.15) is odd and periodic and no such assumptions are made on $f$. Controversy over the solution of the vibrating string continued until 1807 when Fourier presented a paper on heat transfer in which he was able to solve for the steady state temperature $T(x,y)$ of a rectangular metal plate with one edge placed on the interval $[-L,L]$ on the $x$-axis where the initial temperature on the edge $T(x,0) = f(x)$ is known but "arbitrary." Fourier's solution is based on the assumption that an arbitrary function $f$ can be represented as a series of the form

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left( a_n \cos(\frac{n\pi x}{L}) + b_n \sin(\frac{n\pi x}{L}) \right).$$

Moreover, if we take the interval $[0,L]$, then it is sufficient to use only sines or only cosines in the representation as in our discussion above. If we take $L = \pi$, then the Fourier series for $f$ over the interval $[-\pi,\pi]$ is given by

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left( a_n \cos nx + b_n \sin nx \right). \tag{2.16}$$

Fourier argued that if the *Fourier coefficients* $a_0, a_1, \ldots, b_1, b_2, \ldots$ could actually be determined, in other words if the equation (2.16) could be solved, then it must be valid. If we ignore all questions about the convergence of this infinite series, we can compute various coefficients $a_k$ and $b_k$ as in the following. We start with the definition of a periodic function.

**Definition 81.** A function $f : \mathbb{R} \to \mathbb{R}$ is periodic with period $L$ if for all $x$,

$$f(x+L) = f(x).$$

The most important periodic functions are the trigonometric functions $\sin x$ and $\cos x$, each with period $2\pi$. Clearly, the functions $\sin\left(\frac{2\pi x}{L}\right)$ and $\cos\left(\frac{2\pi x}{L}\right)$ are also periodic with period $L$. The importance of these functions is that almost every periodic function is the infinite sum of these trigonometric functions.

## Orthogonality Relations

The theory of Fourier series is based on orthogonality relations for sines and cosines. By analogy with the concept of orthogonal vectors in $\mathbb{R}^n$, two functions $f$ and $g$ are said to be orthogonal over the interval $[a, b]$ if

$$\int_a^b f(x)\, g(x)\, dx = 0.$$

Recalling the trigonometric identities

$$2\cos nx \sin mx = \sin(n+m)x - \sin(n-m)x,$$

$$2\cos nx \cos mx = \cos(n+m)x + \cos(n-m)x,$$

$$2\sin nx \sin mx = \cos(n-m)x - \cos(n+m)x,$$

it is not hard to see the following fundamental orthogonal relations:

$$\int_{-\pi}^{\pi} \cos nx\, dx = \int_{-\pi}^{\pi} \sin nx\, dx = 0, \quad n = 1, 2, \ldots,$$

$$\int_{-\pi}^{\pi} \cos nx \sin mx\, dx = 0, \quad n, m = 1, 2, \ldots,$$

$$\int_{-\pi}^{\pi} \cos nx\, \cos mx\, dx = \int_{-\pi}^{\pi} \sin nx\, \sin mx\, dx = 0, \quad n \neq m,$$

$$\int_{-\pi}^{\pi} \cos^2 nx\, dx = \int_{-\pi}^{\pi} \sin^2 nx\, dx = \pi, \quad n = 1, 2, \ldots.$$

Suppose that for some coefficients $a_k$ and $b_k$ the trigonometric series

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx)$$

converges uniformly in the interval $[-\pi, \pi]$ to a sum $f(x)$. Because we have uniform convergence, the function $f$ is continuous and we can integrate this series term by term. Thus

$$\int_{-\pi}^{\pi} f(x)\, dx = \int_{-\pi}^{\pi} \frac{1}{2} a_0\, dx = \pi a_0.$$

Next, multiply $f(x)$ by $\cos nx$ for some index $n = 1, 2, \ldots$ and integrate term by term to find

$$\begin{aligned} \int_{-\pi}^{\pi} f(x)\, \cos nx\, dx &= \sum_{k=1}^{\infty} a_k \int_{-\pi}^{\pi} \cos kx \cos nx\, dx + \sum_{k=1}^{\infty} b_k \int_{-\pi}^{\pi} \sin kx \cos nx\, dx \\ &= \pi a_n + 0 = \pi a_n. \end{aligned}$$

These all follow from the orthogonality relations described above. Similarly, multiplication by $\sin nx$ and integration gives

$$\int_{-\pi}^{\pi} f(x)\, \sin nx\, dx = \pi b_n.$$

**Definition 82.** The Fourier series of the $2\pi$ periodic function $f : [-\pi, \pi] \to \mathbb{R}$ which is bounded and Riemann integrable on $[-\pi, \pi]$ is given by

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} \left(a_k \cos kx + b_k \sin kx\right),$$

where

$$a_0 = \frac{1}{2} \int_{-\pi}^{\pi} f(x)\, dx$$

and

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x)\, \cos nx\, dx, \quad n = 0, 1, 2, \ldots$$

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x)\, \sin nx\, dx, \quad n = 1, 2, \ldots.$$

The coefficients $a_n$ and $b_n$ are called the Fourier coefficients of $f$ (or amplitudes) for the Fourier series.

Clearly, such a definition can be applied to those functions for which the above integrals exist. The point here, as we will see, is that most functions are actually equal to their Fourier series. Fourier's work, even though considered by his contemporaries to be clumsy and non-rigorous, contains a real innovation. He transformed the existence of a series representation into a geometrically obvious fact that the area bounded by $y = f(x) \sin mx$ and the $x$-axis between $x = -\pi$ and $x = \pi$ can always be computed. In other words he claimed that the Fourier coefficients of an arbitrary bounded function could always be determined.

**Example 87.** In calculating Fourier coefficients of a given function $f$, it is useful to take advantage of special symmetries. For example, if $f$ is an even function, then $f(x) \sin nx$ is odd, then it is clear without calculation that Fourier sine coefficients of $f$ must vanish; in other words $b_n = 0$ for all $n$. Similarly, if $f$ is an odd function, or if it differs by a constant from an odd function, then $a_n = 0$ for all $n \geq 1$ and the formula for $b_n$ may be simplified. Now, we find the Fourier coefficients of the following functions.

a) Let $f(x) = x^2$. Since $f(-x) = f(x)$ this is an even function and its Fourier series reduces to

$$x^2 = \sum_{n=1}^{\infty} a_n \cos nx.$$

To get the Fourier coefficients we consider the following integrals and apply integration by parts to obtain:

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} x^2 \cos nx\, dx = \frac{4}{n^2}(-1)^n \quad n > 0,$$

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{\pi} x^2\, dx = \frac{2\pi^2}{3}.$$

Thus, the function $f(x) = x^2$ in the interval $[-\pi, \pi]$ can be written as

$$x^2 = \frac{\pi^2}{3} + 4\sum_{n=1}^{\infty} \frac{(-1)^n \cos nx}{n^2}.$$

Note that, if we plug the particular point $x = \pi$ into the Fourier series for the function $f(x) = x^2$ we discover a remarkable connection between the irrational number $\pi$ and the integers

$$\pi^2 = \frac{\pi^2}{3} + 4\sum_{n=1}^{\infty} \frac{1}{n^2}$$

which implies

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}.$$

b) Let

$$f(x) = \begin{cases} -1 & \text{if } -\pi < x < 0 \\ 1 & \text{if } 0 < x < \pi; \end{cases}$$

$f$ is sometimes called the *square wave* function. This is an odd function whose Fourier coefficients are $a_n = 0$ and

$$b_n = \frac{2}{\pi}\int_{-\pi}^{\pi} \sin nx dx = \frac{2}{n\pi}[1 - (-1)^n],$$

and therefore

$$b_n = \begin{cases} \frac{4}{n\pi} & \text{if } n \text{ is odd} \\ 0 & \text{if } n \text{ is even.} \end{cases}$$

Except at its points of discontinuity, the above function meets the requirements of convergence in Theorem 108 below; thus we can even write (Figure 2.17)

$$\frac{4}{\pi}\sum_{k=0}^{\infty} \frac{1}{2k+1} \sin(2k+1)x = \begin{cases} -1 & \text{if } -\pi < x < 0 \\ 1 & \text{if } 0 < x < \pi. \end{cases}$$

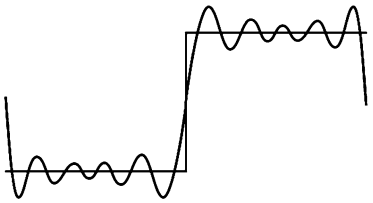


Figure 2.17: Fourier series of the square wave

## Convergence of Fourier Series

It is natural to ask whether a Fourier series of a given continuous periodic function will actually converge to the function, uniformly or even pointwise. If $f$ is smooth, then its Fourier coefficients decay to zero very quickly and it can be shown that the Fourier series converge. The issue becomes much more subtle if $f$ is not smooth. For example there are continuous functions with divergent Fourier series (see [20], p. 219). Thus one has to be careful to specify a precise sense in which the series converge. In fact, most books written on Fourier analysis or harmonic analysis devote a significant portion to this matter.

We start with an important positive development in the rate of convergence for Fourier series by considering the following approximation problem. We say a function $f$ is *square integrable* over $[a, b]$ if it is Riemann integrable over $[a, b]$ and the integral $\int_a^b f(x)^2 dx$ is finite. Now, let

$$s_n(f) = \frac{1}{2}a_0 + \sum_{k=1}^{n} (a_k \cos kx + b_k \sin kx)$$

be the $n$th partial sum of the Fourier series of $f$. Also consider

$$T(x) = \frac{1}{2}c_0 + \sum_{k=1}^{n} (c_k \cos kx + d_k \sin kx)\,. \tag{2.17}$$

Clearly $T(x)$ is a trigonometric polynomial of degree at most $n$. The question here is how should the coefficients $c_k$ and $d_k$ be chosen to minimize the integral

$$\int_{-\pi}^{\pi} [f(x) - T(x)]^2 dx?$$

**Theorem 107** (Mean-square approximation)**.** *The Fourier polynomial $s_n(f)$ gives the best square approximation to $f$ among all trigonometric polynomials of degree less than or equal to $n$.*

*Proof.* Let $T(x)$ be defined as in (2.17) and consider

$$\int_{-\pi}^{\pi} [f(x) - T(x)]^2 dx = \int_{-\pi}^{\pi} f(x)^2\, dx - 2\int_{-\pi}^{\pi} f(x)T(x)\, dx + \int_{-\pi}^{\pi} T(x)^2\, dx.$$

By using the linearity and the orthogonality of the trigonometric system, we can write each of the last two integrals in terms of the Fourier coefficients of $f$ and $T$. Indeed,

$$\begin{aligned}\int_{-\pi}^{\pi} f(x)T(x)dx &= \frac{c_0}{2}\int_{-\pi}^{\pi} f(x)dx + \sum_{k=1}^{n} c_k \int_{-\pi}^{\pi} f(x)\cos kx dx \\ &\quad + \sum_{k=1}^{n} d_k \int_{-\pi}^{\pi} f(x)\sin kx dx \\ &= \pi\left[\frac{c_0 a_0}{2} + \sum_{k=1}^{n}(c_k a_k + d_k b_k)\right]\end{aligned}$$

and

$$\int_{-\pi}^{\pi} T(x)^2\,dx = \pi\left[\frac{c_0^2}{2} + \sum_{k=1}^{n}(c_k^2 + d_k^2)\right].$$

Now, since $c_k^2 - 2c_k a_k = (c_k - a_k)^2 - a_k^2$, we can write:

$$\begin{aligned}\frac{1}{\pi}\int_{-\pi}^{\pi}[f(x) - T(x)]^2 dx &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)^2\,dx - \frac{a_0^2}{2} - \sum_{k=1}^{n}(a_k^2 + b_k^2) \\ &\quad + \frac{(c_0 - a_0)^2}{2} + \sum_{k=1}^{n}\left((c_k - a_k)^2 + (d_k - b_k)^2\right).\end{aligned}$$

The right-hand side is minimized when $c_k = a_k$ and $d_k = b_k$ for all $k$. In other words when $T(x) = s_n(f)$. Note that in this case we have:

$$\begin{aligned}\frac{1}{\pi}\int_{-\pi}^{\pi}[f(x) - s_n(f)]^2 dx &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)^2\,dx - \frac{a_0^2}{2} - \sum_{k=1}^{n}(a_k^2 + b_k^2) \\ &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)^2\,dx - \frac{1}{\pi}\int_{-\pi}^{\pi} s_n(f)(x)^2\,dx.\end{aligned}$$

□

**Corollary 19** (The Weierstrass approximation theorem- trigonometric form)**.** *Let $f(x)$ be continuous function on the interval $[-\pi, \pi]$ and $f(-\pi) = f(\pi)$. Then for each $\epsilon > 0$ there is a trigonometric polynomial $T(x)$ such that*

$$|f(x) - T(x)| < \epsilon \quad \textit{for} \quad -\pi \le x \le \pi.$$

The proof of this theorem is left as an exercise (see Exercise 6 below). Recall that a function $f : [a, b] \to \mathbb{R}$ is called piecewise continuous if $[a, b]$ has a finite partition $a = x_0 < x_1 < \cdots < x_n = b$ such that $f$ is continuous and bounded on each open interval $(x_i, x_{i+1})$. The idea of the proof of the above theorem lies in the following two facts of continuous periodic piecewise linear functions:

- Every continuous periodic function can be approximated uniformly by a continuous piecewise linear function with the same period.

- The Fourier coefficients of a continuous, periodic, piecewise linear function $f$ satisfy $|a_n| \leq \frac{C}{n^2}$ and $|b_n| \leq \frac{C}{n^2}$ for some constant $C > 0$. This shows that the Fourier series converge uniformly to $f(x)$.

Observe that the trigonometric form of the Weierstrass Approximation Theorem is closely related to the algebraic form, as given in the Section 2.4 on approximation by polynomials in this book.

**Remark 57.** Assume $f$ is a real-valued Riemann integrable function and its $L^2$-norm is

$$||f||_2 = \left(\frac{1}{\pi}\int_{-\pi}^{\pi} f(x)^2\,dx\right)^{1/2}.$$

Let $\mathcal{T}_n$ denote the collection of trigonometric polynomials of degree at most $n$. Then we can restate the mean-square theorem above as: the partial sum $s_n(f)$ is the nearest point to $f$ out of $\mathcal{T}_n$, that is

$$\inf_{T\in\mathcal{T}_n} ||f - T||_2 = ||f - s_n(f)||_2.$$

Following Bessel's inequality is valid for more general systems of orthogonal functions. Here our aim is to show that the Fourier coefficients $a_k$ and $b_k$ of square integrable functions tend to zero as $n \to \infty$.

**Corollary 20** (Bessel's inequality)**.** *Let $s_n(f)$ be the nth partial sum of the Fourier series for $f$; then*

$$||s_n(f)||_2 \leq ||f||_2.$$

*Proof.* Note that

$$||f - s_n(f)||_2^2 = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)^2\,dx - \frac{a_0^2}{2} - \sum_{k=1}^{n}(a_k^2 + b_k^2) = ||f||_2^2 - ||s_n(f)||_2^2.$$

Since $||f - s_n(f)||_2 \geq 0$, we have

$$||s_n(f)||_2^2 = \int_{-\pi}^{\pi} s_n(f)(x)^2\,dx = \frac{a_0^2}{2} + \sum_{k=1}^{n}(a_k^2 + b_k^2) \leq \int_{-\pi}^{\pi} f(x)^2\,dx = ||f||_2^2. \quad \square$$

**Remark 58** (Uniform Convergence)**.** The trigonometric polynomials belong to the set of all $2\pi$-periodic continuous functions $f : \mathbb{R} \to \mathbb{R}$, a space we will denote by $C(2\pi)$. If the Fourier series of a function $f \in C(2\pi)$ is uniformly convergent, then the series must actually converge to $f$. Clearly, if a trigonometric series is uniformly convergent then its sum defines a continuous function (recall our discussion of uniform convergence on "modes of convergence" Section 2.3). Let us call this function $g \in C(2\pi)$ in this case. Now, all we need is to observe that $g$ has the same Fourier coefficients as $f$. This is so because if $s_n(f)$ converges uniformly to $g$, then $s_n(f)\cos kx$ converges uniformly to $g(x)\cos kx$. Interchanging the limit and integral, we obtain

$$\frac{1}{\pi}\int_{-\pi}^{\pi} g(x)\cos kx\,dx = \lim_{n\to\infty}\frac{1}{\pi}\int_{-\pi}^{\pi} s_n(f)(x)\cos kx\,dx = a_k.$$

Similarly, $\frac{1}{\pi}\int_{-\pi}^{\pi} g(x)\cos kx\,dx = b_k$.
Moreover, if the Fourier coefficients for $f$ satisfy

$$\sum_n |a_n| < \infty \quad \text{and} \quad \sum_n |b_n| < \infty,$$

then using the well-known Weierstrass M-test (see Theorem 97 in Section 2.3), we can conclude that the Fourier series for $f$ is uniformly convergent in $\mathbb{R}$. Therefore, if we are given that $f \in C(2\pi)$, then by the above discussion it follows that the Fourier series for $f$ converges uniformly to $f$.

In the following we study partial sums of an Fourier series via the Dirichlet kernel. To better understand the pointwise convergence of Fourier series, it would be useful to have an expression for the partial sums $s_n(f)$ not involving a sum; this is done with the Dirichlet formula.

**Definition 83** (Dirichlet Kernel)**.** The function

$$D_n(x) = \frac{\sin(n+\frac{1}{2})x}{2\sin\frac{1}{2}x} = \frac{1}{2} + \sum_{k=1}^{n}\cos kx$$

is known as the *Dirichlet kernel* (Figure 2.18).

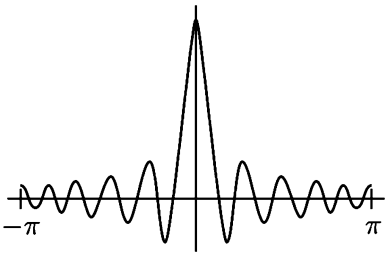



Figure 2.18: The Dirichlet kernel for $n = 10$

First note that the equality of

$$\frac{\sin(n+\frac{1}{2})x}{2\sin\frac{1}{2}x} = \frac{1}{2} + \sum_{k=1}^{n}\cos kx$$

follows from the identity

$$2\sin\frac{1}{2}x\cos kx = \sin\left(k+\frac{1}{2}\right)x - \sin\left(k-\frac{1}{2}\right)x.$$

**Lemma 11.** *The following are the properties of the Dirichlet kernel.*

*a) The Dirichlet kernel is an even function with period $2\pi$, i.e., $D_n(-x) = D_n(x)$ and $D_n(x + 2\pi) = D_n(x)$.*

*b)* $\frac{1}{\pi}\int_{-\pi}^{\pi} D_n(x)\,dx = 1 \quad n = 1, 2, \ldots.$

*c) The nth partial sum $s_n(x)$ of the Fourier series of a function $f$ can be written as*

$$s_n(x) = \frac{1}{\pi}\int_{-\pi}^{\pi} D_n(f)f(x+t)\,dt.$$

*The above formula is called the Dirichlet formula for the nth partial sum of a Fourier series.*

*d)* $\int_{-\pi}^{\pi} |D_n(t)|\,dt \geq C\log n, \quad n \to \infty$ *for some constant* $C > 0$.

*Proof.* a) is clear and b) follows from the simple fact that

$$\int_{-\pi}^{\pi} \cos kx dx = 0 \text{ for } k = 1, 2, \ldots.$$

For c) we write:

$$s_n(x) = \frac{1}{\pi}\int_{-\pi}^{\pi}\left\{\frac{1}{2} + \sum_{k=1}^{n}(\cos kx \cos kt + \sin ka \sin kt)\right\} f(t)dt$$

and therefore,

$$s_n(x) = \frac{1}{\pi}\int_{-\pi}^{\pi}\left\{\frac{1}{2} + \sum_{k=1}^{n}\cos k(x-t)\right\} f(t)dt = \frac{1}{\pi}\int_{-\pi}^{\pi} D_n(x-t)f(t)dt.$$

If $f$ is $2\pi$-periodic, then we can also write

$$s_n(x) = \frac{1}{\pi}\int_{-\pi}^{\pi} D_n(t)f(x+t)dt$$

and moreover

$$s_n(x) - f(x) = \frac{1}{\pi}\int_{-\pi}^{\pi} D_n(t)[f(x+t) - f(x)]dt.$$

For part d) we refer the reader to [49], p. 50. □

In light of the Dirichlet formula we see that in order to prove $s_n(x) \to f(x)$ as $n \to \infty$, we only need to show that the above integral tends to zero. However, $D_n(t)$ alternates sign in $[-\pi, \pi]$ and the property d) in the above lemma makes it clear that we need a better kernel; this "good" kernel is called the Fejér kernel for approximation, as we shall discuss soon. Nevertheless, one can find extra conditions on the function $f$ under which $s_n \to f$ uniformly (see [49], p. 41).

Even though the Fourier coefficients $a_k$ and $b_k$ are determined by the values of the function $f$ throughout the interval $[-\pi, \pi]$, the convergence of the resulting Fourier series at any point $x$ depends on the behavior of $f$ in an arbitrarily small neighborhood of $x$ as illustrated in the following theorem.

**Theorem 108** ([20], p. 205)**.** *Let $f$ be square integrable over $[-\pi, \pi]$, and let $s_n$ be the nth partial sum of its Fourier series. Let $f$ be extended to the real line with $f(x + 2\pi) = f(x)$. Suppose at some point $x \in \mathbb{R}$ the extended function satisfies the condition*

$$|f(x + t) - f(x)| \leq C|t|, \quad \textit{where} \quad |t| < \delta,$$

*for some positive constants $C$ and $\delta$. Then $s_n(x)$ converges to $f(x)$ when $n$ approaches $\infty$.*

To motivate the definition of the Fejér kernel and a result on uniform convergence called the Fejér theorem, we begin with a simple observation that "averaging preserves convergence." Suppose we are given a sequence of real numbers $\{s_n\}$ and consider the sequence formed by their arithmetic means (or Cesàro sums)

$$\sigma_n = \frac{s_1 + s_2 + \cdots + s_n}{n}.$$

**Proposition 32.** *If $s_n \to s$, then $\sigma_n \to s$.*

*Proof.* The idea of the proof depends on the fact that if $\{s_n\}$ is convergent, then given $\epsilon > 0$, we can choose $n$ such that $|s_k - s| < \epsilon$ for all $k \geq n$. On the other hand if $\{s_n\}$ is convergent, then it is bounded; thus there is an $M > 0$ such that $|s_n| \leq M$ for all $n$. Fixing this $n$, one can rewrite $\sigma_N$ for $N > n$ as

$$\sigma_N = \frac{s_1 + \cdots + s_n}{N} + \frac{s_{n+1} + \cdots + s_N}{N}$$

and prove that $s - 2\epsilon < \sigma_N < s + 2\epsilon$. □

Next, we take arithmetic means of the Dirichlet kernel to obtain the Fejér kernel.

**Definition 84.** The function

$$K_n(x) = \frac{1}{n+1} \sum_{k=0}^{n} D_k(x) = \frac{1}{n+1} \frac{\sin^2 \frac{1}{2}(n+1)x}{2 \sin^2 \frac{1}{2}x}$$

is known as the *Fejér kernel*.

Following figure shows the Fejér kernel when $n = 10$ (Figure 2.19).

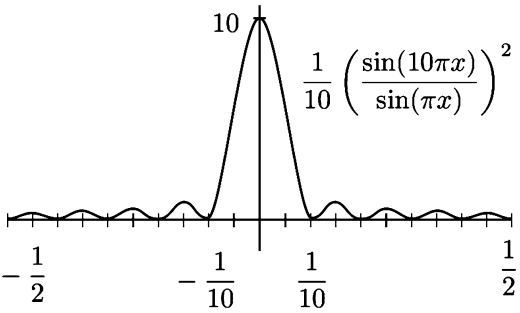



Figure 2.19: The Fejér kernel for $n = 10$

Note that $K_n$ is an even, non-negative trigonometric polynomial and one of the property of the Dirichlet kernel (see Lemma 11, part b)) implies that

$$\frac{1}{\pi}\int_{-\pi}^{\pi} K_n(x)dx = 1, \quad \text{for} \quad n = 1, 2, \ldots.$$

It is not difficult to see that $\sigma_n(f)$ is a good approximation to $f$ in the desirable situation of $L_2$ norm. In other words, $||f - \sigma_n(f)(x)||_2 \to 0$ as $n \to \infty$ (see Exercise 7 below). The point of the next theorem is that $\sigma_n(f)(x)$ is actually a good *uniform* approximation to $f$. We state the Fejér theorem without proof. Its proof could be accessed in most books on Fourier Analysis. For example one can consult [49], p. 53.

**Theorem 109** (Fejér theorem)**.** *Let $f(x)$ be continuous on $\mathbb{R}$ and periodic with period $2\pi$. Let $\sigma_n(x) = \sigma_n(f)(x) = \dfrac{1}{n+1}\displaystyle\sum_{k=0}^{n} s_k(x)$ be the arithmetic means of the partial sums of the Fourier series of $f$. Then*

$$\sigma_n(x) \to f(x) \ \textit{uniformly on } \mathbb{R} \textit{ as } n \to \infty.$$

## Complex Fourier Series and Fourier Transform

It is convenient to express the Fourier transform in the notation of complex exponentials

$$e^{it} = \cos t + i \sin t.$$

For a $2\pi$ periodic function $f : \mathbb{R} \to \mathbb{C}$ we define the complex form Fourier series of $f$ by

$$f(t) = \sum_{k=-\infty}^{\infty} c_k e^{ikt},$$

where now we have a single formula for $c_k$:

$$c_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{-ikt}dt \qquad n = 0, \pm 1, \pm 2, \ldots.$$

Note that $c_0$ is real, i.e., $c_0 = c_{-0} = \overline{c_0}$ , and $c_{-k} = \overline{c_k}$ for $k = 0, 1, \ldots$, where by $\overline{c_k}$ we mean the complex conjugate of $c_k$. The integral of a complex-valued function is defined in terms of the real and imaginary parts, thus we assume that both $\mathrm{Re}f$ and $\mathrm{Im}f$ are integrable on $[-\pi, \pi]$. Namely, $\int f = \int \mathrm{Re}f + \int \mathrm{Im}f$. There are some advantages to this approach; one of which is that complex exponentials form a basis for $L^2$. The exponentials $e^{ikt}$ are now an orthonormal set. Specifically, we define

$$\langle f, g \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t)\overline{g(t)}\, dt,$$

and then we have

$$\langle e^{int}, e^{imt} \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{int} e^{-imt} dt = \begin{cases} 0 & \text{for } m \neq n \\ 1 & \text{for } m = n. \end{cases}$$

Furthermore, with this choice we can define the $L^2$-norm by

$$||f||_2 = \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(t)|^2\, dt \right)^{1/2}.$$

The Dirichlet and Fejér kernels are essentially the same in this case, except now we take $s_n(f)(x) = \sum_{k=-n}^{n} c_k e^{ikx}$.

Loosely speaking, the Fourier transform and its inversion are the limiting case of Fourier series or the continuous analog of Fourier series. Even though some ideas carry over between Fourier series and Fourier transforms, some do not. Fourier transforms are considered to be a simpler approach and have other advantages. In what follows we will be dealing with a complex-valued function $f(t)$ defined on the real line which is Riemann integrable over each bounded interval. For $p > 0$, we say $f \in L^p$ when $|f(t)|^p$ is integrable over $\mathbb{R}$.

**Definition 85.** Let $f$ be integrable, in other words $\int_{-\infty}^{\infty} |f(x)| dx < \infty$. The Fourier transform of $f(x)$ is denoted by $\widehat{f}(x)$ and defined as:

$$\widehat{f}(x) = \int_{-\infty}^{\infty} f(t)\, e^{-ixt}\, dt \qquad -\infty < x < \infty.$$

The Fourier transform of an integrable function $f$ plays the role of the sequence of Fourier coefficients and the function can be recovered from its Fourier transform by the formula

$$f(t) = \int_{-\infty}^{\infty} \widehat{f}(t)\, e^{ixt}\, dt \qquad -\infty < x < \infty,$$

which is a continuous analog of a Fourier series expansion. The following are some simple properties:

**Lemma 12.** *Suppose that $f$ is integrable, then*

*a)* $|\widehat{f}(x)| < \infty$.

*b) The Fourier transform is a linear operator.*

*Proof.* a) $|\widehat{f}(x)| \leq \int_{-\infty}^{\infty} |f(t)|\, dt < \infty$ provided that $f$ is integrable.

b) If $f$ and $g$ are integrable functions and $\alpha, \beta \in \mathbb{C}$ then $\widehat{\alpha f + \beta} g = \alpha \widehat{f} + \beta \widehat{g}$ follows directly from the definition of Fourier transform. □

**Example 88.** Consider the function

$$f(t) = \begin{cases} e^{-ct} & \text{for } t \geq 0 \\ 0 & \text{for } t < 0 \end{cases}$$

for some constant $c > 0$. Then the Fourier transform of $f$ is

$$\widehat{f}(x) = \int_0^{\infty} e^{-ixt}\, e^{-ct}\, dt = \int_0^{\infty} e^{-ct}(\cos xt - i \sin xt) dt = \frac{c - ix}{c^2 + x^2}.$$

Observe that $\widehat{f}$ is not an integrable function.

**Definition 86.** The *convolution* of two functions $f(x)$ and $g(x)$ in $L^1$ is defined by the integral

$$(f * g)(x) = \int_{\mathbb{R}} f(y) g(x - y) dy.$$

By changing variables $y \mapsto x - y$, we may also write

$$(f * g)(x) = \int_{\mathbb{R}} f(x - y) g(y) dy$$

so that $f * g = g * f$. It is easy to see that the integral converges absolutely if one of the functions $f$ or $g$ is bounded. The most important property of the Fourier transform is that it converts convolution to ordinary pointwise multiplication:

$$\widehat{(f * g)}(x) = \widehat{f}(x) \widehat{g}(x).$$

For a proof, one can change the order of integration by using Fubini's Theorem (for Fubini's theorem see, e.g., [50], pp. 75–80) to write

$$\begin{aligned} \widehat{(f * g)}(x) &= \int_{-\infty}^{\infty} e^{-ixt} \int_{-\infty}^{\infty} f(s) g(t - s) ds\, dt \\ &= \int_{-\infty}^{\infty} e^{-ixs} f(s) \int_{-\infty}^{\infty} e^{-ix(t-s)}\, g(t - s) dt\, ds \\ &= \widehat{f}(x) \widehat{g}(x). \end{aligned}$$

This relationship is important when one tries to solve partial differential equations. For example, it could be the case that one has the product of two Fourier transforms; by the above relationship we can recognize it as a Fourier transform of a single function formed by the convolution.

## Fourier Transform and the Laplacian

For applications to the theory of partial differential equations, one of the important Fourier transforms is the one defined on Euclidean space $\mathbb{R}^n$. Suppose $f : \mathbb{R}^n \to \mathbb{C}$ is the *plane wave function* defined as:

$$f(x) = c_\xi e^{2\pi i x\cdot\xi}.$$

Here both $x, \xi \in \mathbb{R}^n$, where $x$ is the position, $\xi$ is the frequency of the plane wave, $x \cdot \xi$ is the dot product, and $c_\xi$ is a complex number whose magnitude is the *amplitude*. The reason why this function is named the plane wave function stems from the fact that if one considers the hyperplane $H_\alpha = \{x : \ x \cdot \xi = \alpha\}$, then $f(x)$ is constant on $H_\alpha$ and the value taken by $f$ on $H_\alpha$ is equal to the value taken on $H_{\alpha+2\pi}$. It turns out that if $f$ is sufficiently smooth, then it can be represented as the superposition of plane waves. The following formulas show how to determine a Fourier transformed function from the original function and vice versa. Namely, we have:

$$f(x) = \int_{\mathbb{R}^n} \widehat{f}(\xi) e^{2\pi i x\cdot\xi} d\xi,$$

and the Fourier transform is

$$\widehat{f}(\xi) = \int_{\mathbb{R}^n} f(x) e^{-2\pi i x\cdot\xi} dx.$$

Again, it turns out that if $f$ is sufficiently smooth, one can easily verify convergence of these integrals. Now given a function $f : \mathbb{R}^n \to \mathbb{C}$, its *Laplacian* $\triangle f$ is defined by the formula

$$\triangle f(x) = \sum_{i=1}^{n} \frac{\partial^2 f}{\partial x_i^2},$$

where $f$ is a function of $n$ real variables. For simplicity, consider only functions $f$ where the above formula is well defined. Now suppose $f$ is a plane wave function $f(x) = e^{2\pi i x\cdot\xi}$, then its Laplacian is

$$\triangle\, e^{2\pi i x\cdot\xi} = -4\pi^2|\xi|^2 e^{2\pi i x\cdot\xi}.$$

Thus, the effect of the Laplacian on the plane wave function is to multiply by the scalar $-4\pi^2|\xi|^2$. This means that plane wave function is an eigenfunction for the Laplacian $\triangle$. Even though in general there is no obvious relationship between a function $f$ and its Laplacian $\triangle f$, we have a very different situation for the wave function. If we study the Laplacian via the Fourier transform, then we can write an arbitrary function as a superposition of plane waves and the Laplacian has a striking effect on each plane wave. We can even obtain a formula for the Laplacian of the general function. More precisely, if we have a nice enough function $f$ so that we can interchange the integral $\int$ and the Laplacian $\triangle$,

$$\triangle f(x) = \triangle \int_{\mathbb{R}^n} \hat{f}(\xi) e^{2\pi i x\cdot\xi} d\xi = \int_{\mathbb{R}^n} \hat{f}(\xi) \triangle\, e^{2\pi i x\cdot\xi} d\xi.$$

Then the Laplacian

$$\triangle f(x) = \int_{\mathbb{R}^n} (-4\pi^2|\xi|^2)\hat{f}(\xi)e^{2\pi ix\cdot\xi}d\xi.$$

This formula represents $\triangle f(x)$ as a superposition of plane waves. Moreover, such a representation is unique. Recalling the Fourier inversion formula

$$\triangle f(x) = \int_{\mathbb{R}^n} \widehat{\triangle f}(\xi)e^{2\pi ix\cdot\xi}d\xi,$$

or using integration by parts formula in the definition of Fourier transform, we obtain

$$\widehat{\triangle f}(\xi) = (-4\pi^2|\xi|^2)\widehat{f}(\xi).$$

This identity shows that the Fourier transform diagonalizes the Laplacian operator, i.e., the operation of taking the Laplacian is multiplication of a function $F(\xi)$ by the multiplier $(-4\pi^2|\xi|^2)$.

**Remark 59.** The classical text in Fourier Series has been Zygmund's *Trigonometric Series* [56] which was published first in 1930. Stein and Sakarachi's book *Fourier Analysis an Introduction* [49], Folland's *Fourier Series and its Applications* [21], and Körner's *Fourier Analysis* [33] are books that all provide a solid introduction to the topic. Since Fourier analysis has applications to a wide range of areas, there is an ample supply of books on applications. For example, B. Hubbard's [27] *The World According to Wavelets*, *Fourier Series and Numerical Methods for Partial Differential Equations* by R. Bernatz [10], *Fourier Analysis in Probability Theory* by T. Kawata [31], and *Fourier Analysis of Time Series* by Bloomfield [12] are good sources.

## Exercises

1. Let $f : \mathbb{R} \to \mathbb{R}$ be a $2\pi$-periodic and Riemann integrable function on $[-\pi, \pi]$. If $f$ is even (respectively odd) show that the Fourier series can be written using only cosine (respectively sine) terms.

2. Let $f : \mathbb{R} \to \mathbb{R}$ be a unique function such that $f(x) = x$ if $\pi \leq x < \pi$ and $f(x + 2n\pi) = f(x)$ for all $n \in \mathbb{Z}$.

   a) Show that Fourier series for $f$ is $\displaystyle\sum_{n=1}^{\infty} \frac{(-1)^{n+1}2\sin nx}{n}$.

   b) Prove that the series above does not converge uniformly.

3. Define $f(x) = \pi - x$ for $0 < x < 2\pi$, where $f(0) = f(2\pi) = 0$. Extend $f$ to a $2\pi$-periodic function on $\mathbb{R}$. Prove that the Fourier series for $f$ is $2\displaystyle\sum_{n=1}^{\infty} \frac{\sin nx}{n}$.

4. On the vector space

$$L^2[-\pi,\pi] = \left\{ f : [-\pi,\pi] \to \mathbb{C} : \quad \int_{-\pi}^{\pi} |f|^2 < \infty \right\},$$

show that

$$\langle f, g \rangle = \int_{-\pi}^{\pi} f(x)\,\overline{g(x)}\,dx$$

is an inner product.

5. Let $f(x) = |x|$ for $x \in [-\pi, \pi]$ be extended periodically to the whole line by $f(x + 2\pi) = f(x)$. Find Fourier coefficients and the Fourier series for $f$.

6. If $f$ and $g$ are Riemann integrable on $[-\pi, \pi]$, then show that $s_n(\alpha f + \beta g) = \alpha s_n(f) + \beta s_n(g)$ for every $n$. In other words, show that the map $f \mapsto s_n(f)$ is linear.

7. Prove the trigonometric form of the Weierstrass approximation theorem given in Corollary 19.

8. Prove that $||f - \sigma_n(f)(x)||_2 \to 0$ as $n \to \infty$, where $\sigma_n(f)(x) = \dfrac{1}{n+1} \displaystyle\sum_{k=0}^{n} s_k(x)$.

9. Let $f(t) = e^{-c|t|}$ for some constant $c > 0$. Show that its Fourier transform is $\widehat{f}(x) = \dfrac{2c}{c^2 + x^2}$.

10. Let $a \neq 0$ and $b$ be real constants and $g(t) = f(at + b)$. Prove that

$$\widehat{g}(x) = \frac{1}{|a|} e^{\frac{ibx}{a}} \widehat{f}\left(\frac{x}{a}\right).$$

11. Show that the operation of convolution behaves much like multiplication of numbers: it is commutative, associative, and distributive over addition:

$$f * g = g * f, \qquad f * (g * h) = (f * g) * h, \qquad f * (g + h) = f * g + f * h.$$

12. Consider the convolution

$$(f * g)(x) = \int_{\mathbb{R}} f(x - y) g(y) dy.$$

a) Show that $f * g$ is uniformly continuous when $f$ is integrable and $g$ is bounded.

b) If in addition $g$ is integrable, show that $(f * g)(x) \to 0$ as $|x| \to \infty$.

13. Let $f$ be twice differentiable real-valued function on $[0, 2\pi]$ with

$$\int_0^{2\pi} f(x)\,dx = 0 = f(2\pi) - f(0).$$

Show that

$$\int_0^{2\pi} (f(x))^2\,dx \leq \int_0^{2\pi} (f'(x))^2\,dx.$$

Hint: Consider Fourier series for $f$ and use Parseval's Identity (see [26], p. 577).

## 2.7 Lebesgue Measure and Integration

The treatment of integration developed by French mathematician Henri Lebesgue (1875–1941) has become indispensable in many areas of mathematics because it not only allows us to integrate many more functions but also enables us to understand when the integral of the limit functions is equal to the limit of the integrals. In his thesis in 1902, Lebesgue presented an extension of the Riemann integral. The Lebesgue integral is defined for what are called *measurable* functions, a class that includes the Riemann integrable functions. Moreover, Lebesgue's ideas gave new insights on the differentiability of monotone functions and led to an extension of the fundamental theorem of calculus.

The Riemann integral is surely one of the gems of calculus and works for many functions. Unfortunately it has some shortcomings. Before we list few of these difficulties below, let us recall how we defined the Riemann integral.

**Definition 87.** Let $f$ be a function defined on a bounded interval $[a, b]$. A *partition* $P$ of $[a, b]$ is a finite set of points in $[a, b]$ such that

$$a = x_0 < x_1 < x_2 < \cdots < x_n = b.$$

For $k = 1, 2, \ldots, n$, let

$$M_k = \sup\{f(x) :\ x_{k-1} \leq x \leq x_k\} \quad \text{and} \quad m_k = \inf\{f(x) :\ x_{k-1} \leq x \leq x_k\};$$

then by the *upper sum*, denoted by $U(f, P)$, and the *lower sum*, denoted by $L(f, P)$ we mean:

$$U(f, P) = \sum_{k=1}^{n} M_k(x_k - x_{k-1}) \quad \text{and} \quad L(f, P) = \sum_{k=1}^{n} m_k(x_k - x_{k-1}).$$

Clearly $L(f, P) \leq U(f, P)$ for a particular partition $P$, and it helps to think of upper sums as an overestimate and lower sums as an underestimate for the value of the integral. If $P$ and $Q$ are two partitions of $[a, b]$ with $P \subset Q$, then $Q$ is called a *refinement* of $P$. It can be seen that if $P \subset Q$ then

$$L(f, P) \leq L(f, Q) \quad \text{and} \quad U(f, P) \geq U(f, Q).$$

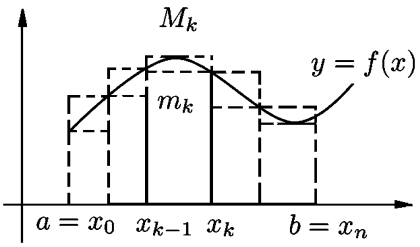



Figure 2.20: Upper and lower sums

Let $\mathcal{P}$ be the collection of all possible partitions of the interval $[a, b]$, and define (Figure 2.20)

$$U(f) = \inf\{U(f, P) : \ P \in \mathcal{P}\} \quad \text{and} \quad L(f) = \sup\{L(f, P) : \ P \in \mathcal{P}\}.$$

Let $f$ be a bounded function on $[a, b]$ and $P$ be any partition of $[a, b]$. First, observe that $U(f, P) \geq L(f, P)$; then one can prove that for any bounded function $f$ on $[a, b]$ we always have $L(f) \leq U(f)$.

**Definition 88.** A bounded function $f$ on $[a, b]$ is *Riemann integrable* over $[a, b]$ if $U(f) = L(f)$ and its integral is defined to be the common value, written as

$$\int_a^b f(x)\, dx = U(f) = L(f).$$

We can also say that $f$ is Riemann integrable if and only if for each $\epsilon > 0$, there exists a partition $P$ such that $U(f, P) - L(f, P) < \epsilon$, that is, its integrability is equivalent to the existence of partitions whose upper and lower sums are arbitrarily close together. More precisely:

**Proposition 33** (Riemann's Criterion)**.** *A bounded real-valued function $f$ on $[a, b]$ is Riemann integrable if and only if for each $\epsilon > 0$ there is a partition $P_\epsilon$ such that*

$$U(f, P_\epsilon) - L(f, P_\epsilon) < \epsilon.$$

Now we give some reasons why the Riemann integral is not good enough. First of all the Riemann integral does not handle functions with many discontinuities as illustrated in the following example.

**Example 89.** Consider Dirichlet's function $f : [0, 1] \to \mathbb{R}$ given by

$$f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \notin \mathbb{Q} \end{cases}.$$

We have seen that this function is discontinuous at every point of $[0, 1]$. Now we turn our attention to integrability. If $P$ is some partition of $[a, b]$, then the density of rationals in $\mathbb{R}$ implies that every subinterval of $P$ contains points for which $f(x) = 1$, and thus $U(f, P) = 1$. Similarly, $L(f, P) = 0$ because irrational

numbers are also dense in $\mathbb{R}$. Since this is the case for every possible partition, $U(f) = 1$ and $L(f) = 0$; the two are not equal; therefore, Dirichlet's function is not Riemann integrable. This example illustrates that not every bounded function is Riemann integrable.

The following is another example showing that Riemann integral does not handle unbounded functions.

**Example 90.** Consider

$$f(x) = \begin{cases} \frac{1}{\sqrt{x}} & \text{if } 0 < x \leq 1 \\ 0 & \text{if } x = 0 \end{cases}.$$

If $x_0 < x_1 < x_2 < \cdots < x_n$ is a partition of $[0, 1]$, then $\sup_{[x_0, x_1]} f = \infty$; thus if we try to find the upper Riemann sum for $f$, we will get $U(f, P) = \infty$. On the other hand, using calculus in order to find the area under the graph of $f$ we set

$$\int_0^1 f(x)\, dx = \lim_{t \downarrow 0} \int_t^1 \frac{1}{\sqrt{x}}\, dx = \lim_{t \downarrow 0} (2\sqrt{x})\,|_t^1 = 2;$$

thus the area under $f$ is 2 not $\infty$.

The Riemann integral does not work well with pointwise limits either. When a sequence of Riemann integrable functions $f_n(x)$ converge pointwise to $f(x)$ on an interval $[a, b]$, the limit function $f$ need not be Riemann integrable. It is also possible that

$$\lim_{n \to \infty} \int_a^b f_n(x)\, dx \neq \int_a^b \lim_{n \to \infty} f_n(x)\, dx,$$

in other words, we cannot interchange the limit and integral.

**Example 91.** Suppose we have a sequence of functions $\{f_n\}$ where the graph of $f_n(x)$ looks like Figure 2.21:

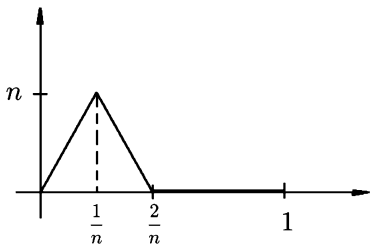



Figure 2.21: Graph of $f_n$

Then, it is easy to see that the sequence $\{f_n(x)\}$ converges pointwise to $f(x) = 0$, yet

$$\int_0^1 f_n(x)\,dx = 1, \text{ for every n, whereas } \int_0^1 f(x)\,dx = 0.$$

The Lebesgue approach of integration reduces the problem of integration to that of the integral of the characteristic function $\chi_E$ of a set $E$, where by the characteristic function of $E$ we mean:

$$\chi_E(x) = \begin{cases} 1 & \text{if } x \in E \\ 0 & \text{if } x \notin E \end{cases}.$$

Lebesgue finds a suitable way to define

$$m(E) = \int_a^b \chi_E\,dx.$$

By $m(E)$ we mean the measure of the set $E$ which we will define later. The point here is that the problem of integration connects with the problem of measure. That is why we first define measure of a set, then introduce measurable functions and then present Lebesgue integral and its fantastic convergence theorems.

## Lebesgue Outer Measure

Let $E \subset \mathbb{R}$, and suppose we want to assign a non-negative number $m(E)$, called the *measure of* $E$, in such a way that the following properties are satisfied:

a) $m([0,1]) = 1$.

b) $m(E + h) = m(E) = m(-E)$, where $E + h = \{x + h : x \in E\}$ and $-E = \{-x : x \in E\}$. That is, congruent sets have the same measure.

c) If $(E_n)$ is any sequence, finite or infinite, of pairwise disjoint sets, then

$$m\left(\bigcup_{n\geq 1} E_n\right) = \sum_{n\geq 1} m(E_n).$$

The above three conditions together will imply that the measure of an interval is simply its length. These three conditions are not only independent but if one assumes the Axiom of Choice they became inconsistent (see the Section 2.8 about the *Banach–Tarski Paradox*). If we look at the last condition and think of the geometric notions of length, area, and volume which only require for those measures to be additive across finitely many disjoint objects, we will see why Lebesgue departed from the last condition above and considered *countably additive* measures.

The notion of *outer measure* is the first step toward an understanding the notion of length. Let $I$ be an open interval of $\mathbb{R}$; we write $|I|$ to denote the length of $I$ which is defined as

$$|I| = \begin{cases} b-a & \text{if } I=(a,b) \\ 0 & \text{if } I=\emptyset \\ \infty & \text{if } I=(-\infty,a) \text{ or } I=(a,\infty) \\ \infty & \text{if } I=(-\infty,\infty). \end{cases}$$

**Definition 89.** Let $A$ be a subset of $\mathbb{R}$. The *outer measure* of $A$, denoted by $m_*(A)$, is defined as

$$m_*(A) = \inf\left\{\sum_{k=1}^{\infty}|I_k| : \{I_k\} \text{ sequence of open intervals such that } A\subset\bigcup_{k=1}^{\infty} I_k\right\},$$

where the infimum is taken over all countable coverings of $A$.

The outer measure is always non-negative but could be infinite, so that in general we have $0 \le m_*(A) \le \infty$. Furthermore from the properties of infimum, for every $\epsilon > 0$, there exists a covering $A \subset \bigcup_{k=1}^{\infty} I_k$ with

$$\sum_{k=1}^{\infty}|I_k| \le m_*(A) + \epsilon.$$

**Properties of the Outer Measure:**

**Theorem 110.**

*a) Every countable subset of $\mathbb{R}$ has outer measure zero.*

*b) $m_*(A) = |A|$ if $A$ is a closed interval.*

*c) $A \subset B$ implies $m_*(A) \le m_*(B)$ (Monotonicity).*

*d) Suppose $t \in \mathbb{R}$ and $A \subset \mathbb{R}$. Then $m_*(A) = m_*(A+t)$ (Translation invariance).*

*e) If $A = \bigcup_{j=1}^{\infty} A_j$, then $m_*(A) \le \sum_{j=1}^{\infty} m_*(A_j)$ (Countable sub-additivity).*

*Proof.* For part a). Suppose $A$ is countably many points and let $A = \{x_1, x_2, \dots\}$ be an enumeration of the points of $A$. Given $\epsilon > 0$, set $I_k$ to be an open interval that contains the point $x_k$ with length $\frac{\epsilon}{2^k}$. The collection $\{I_k\}$ covers $A$ and $\sum |I_k| = \sum \frac{\epsilon}{2^k} = \epsilon$; thus $m_*(A) \le \epsilon$.
For part b). Suppose $A = [a,b]$. Given $\epsilon > 0$, one can set $I_1 = (a-\epsilon/2,\ b+\epsilon/2)$

and for $k > 1$ set $I_k = \emptyset$. Thus, $m_*(A) \leq b - a + \epsilon$. Because this inequality holds for every $\epsilon > 0$, we conclude that $m^*(A) \leq b - a$. For the converse inequality suppose $\{I_k\}$ to be an open cover for $A$. Since $A$ is compact, there is a finite subcover, $A \subset \bigcup_{j=1}^{N} I_j$. Renumbering we get points

$$a = a_1 < \cdots < a_{n+1} = b, \qquad (a_j, a_{j+1}) \subseteq I_j.$$

Then

$$|A| = b - a = a_{n+1} - a_1 \leq \sum_{j=1}^{n} |I_j| \quad \text{thus} \quad |A| \leq m_*(A).$$

Monotonicity and translation invariance properties follows from the definition of outer measure.

For part e) (Countable sub-additivity). We may assume that $m_*(A_j) < \infty$. If $A = \bigcup_{j=1}^{\infty} A_j$, then for any $\epsilon > 0$, the definition of outer measure yields for each $j$ a covering $A_j \subset \bigcup_{k=1}^{\infty} I_{k,j}$ with

$$\sum_{k=1}^{\infty} |I_{k,j}| \leq m_*(A_j) + \frac{\epsilon}{2^j}.$$

Thus, $A \subset \bigcup_{j,k=1}^{\infty} I_{k,j}$ is a covering for $A$, and therefore

$$m_*(A) \leq \sum_{j,k} |I_{k,j}| = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} |I_{k,j}| \leq \sum_{j=1}^{\infty} \left( m_*(A_j) + \frac{\epsilon}{2^j} \right) = \sum_{j=1}^{\infty} m_*(A) + \epsilon$$

holds for every $\epsilon > 0$; whence the countable sub-additivity is proved. □

**Corollary 21.** *Every interval $B$ in $\mathbb{R}$ that contains at least two distinct points $a, b \in \mathbb{R}$ with $a < b$ is uncountable.*

*Proof.* The monotonicity property of outer measure implies $|B| \geq |[a, b]| = b - a$ and because every countable subset of $\mathbb{R}$ has outer measure 0 (see property a)), we can conclude that $B$ is uncountable. □

Suppose $A$ and $B$ are subsets of $\mathbb{R}$ such that $A \cap B = \emptyset$. Do we have

$$m_*(A \cup B) = m_*(A) + m_*(B)?$$

This important property is not in the above list of properties of the outer measure for a good reason. It holds when the sets are not "irregular" but are measurable. Rather than modify $m_*$, we restrict ourselves to the class of "measurable" subsets that we work with.

## Measurable Sets

The notion of measurability isolates a collection of subsets of $\mathbb{R}$ for which the outer measure satisfies good properties, including additivity, and in fact countable additivity, for the disjoint union of sets.

**Definition 90.** A subset $A \subset \mathbb{R}$ is called *measurable* or *Lebesgue measurable* if for every $\epsilon > 0$ there exists an open set $O$ such that $A \subset O$ and

$$m_*(O \setminus A) \leq \epsilon.$$

In other words, Lebesgue measurable sets are defined to be the sets that can be contained "nicely" in open sets such that the difference or the error one is making has a small outer measure. Immediately from the definition we find that $\emptyset$ and $\mathbb{R}$ are measurable. Furthermore, every open set in $\mathbb{R}$ is measurable and if $m_*(A) = 0$ then $A$ is measurable.

### Fundamental Properties of Measurable Sets

Suppose $A$ is a measurable set, then its Lebesgue measure is defined to be its outer measure, denoted as $m(A)$,

$$m_*(A) = m(A) \qquad \text{if } A \text{ is measurable.}$$

Clearly the Lebesgue measure inherits all the properties of the outer measure. Furthermore, collection of measurable sets behave well under the many operations of set theory, countable unions, countable intersections, and complements. In the following we give some fundamental properties of measurable sets. For the proof of these facts consult, e.g., [50], pp. 16–22.

a) Complements, countable unions, and countable intersections of measurable sets are measurable.

b) Any interval is measurable and its measure is its length.

c) (Countable additivity). If $(A_n)_{n=1}^{\infty}$ is a sequence of pairwise disjoint measurable sets then
$$m\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} m(A_n).$$

d) (Continuity). Suppose that $(A_n)_{n=1}^{\infty}$ and $(B_n)_{n=1}^{\infty}$ are countable collection of subsets of $\mathbb{R}$ such that

$$A_1 \supset A_2 \supset \cdots \qquad \text{and} \quad B_1 \subset B_2 \subset \cdots$$

and suppose that $m(A_1) < \infty$. Then

$$m\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} m(A_n), \qquad m\left(\bigcup_{n=1}^{\infty} B_n\right) = \lim_{n\to\infty} m(B_n).$$

e) (Approximation). Suppose $A$ is a measurable subset of $\mathbb{R}$. Then for every $\epsilon > 0$ there exists a closed set $C \subset A$ with

$$m(A \setminus C) \leq \epsilon.$$

If $m(A)$ is finite, there exists a compact set $K$ with $K \subset A$ and

$$m(A \setminus K) \leq \epsilon.$$

**Remark 60.** Are all subsets of $\mathbb{R}$ measurable? The answer to this question is given in the section on the Banach–Tarski paradox (see Section 2.8), where we present a construction of a subset of $\mathbb{R}$ which is not measurable.

## Measurable Functions

If $f : \mathbb{R} \to \mathbb{R}$, recall that $f$ is continuous if and only if $f^{-1}(A)$ is open whenever $A$ is open. Here the inverse image $f^{-1}(A)$ of $A$ is defined as

$$f^{-1}(A) = \{x \in \mathbb{R} : f(x) \in A\}.$$

In this section we will define measurable functions which are "almost" continuous, but in a somewhat weaker sense than that of Riemann integrable functions. That is what we expect, since we want the class of Lebesgue integrable functions to be larger than the class of Riemann integrable functions. Note that we say $f$ is finite valued if $-\infty < f(x) < \infty$ for all $x$. In many applications and in the theory, a function takes on infinite values on at most a set of measure zero.

**Definition 91.** A real-valued function $f$ defined on $\mathbb{R}$ is a *measurable function* if for each open set $A$ of $\mathbb{R}$, the inverse image $f^{-1}(A)$ is measurable.

Note that this property also applies to extended-valued functions if we make the additional hypothesis that both $f^{-1}(\infty)$ and $f^{-1}(-\infty)$ are measurable sets. Also if $f$ is a continuous function with a measurable domain then $f$ is measurable, since open sets are measurable. Moreover monotone functions, step functions, and semicontinuous functions defined on some interval in $\mathbb{R}$ are all measurable. In particular if we think of $A = [-\infty, a)$,

$$f^{-1}([-\infty, a)) = \{x \in \mathbb{R} : \ f < a\}$$

and write $\{f < a\}$ for simplicity instead of $\{x \in \mathbb{R} : \ f < a\}$, we see that there are a number of equivalent statements to the above definition. More precisely, each of the following conditions imply the other three:

a) $\{x : f(x) > a\}$ is measurable for all $a \in \mathbb{R}$.

b) $\{x : f(x) \geq a\}$ is measurable for all $a \in \mathbb{R}$.

c) $\{x : f(x) < a\}$ is measurable for all $a \in \mathbb{R}$.

d) $\{x : f(x) \leq a\}$ is measurable for all $a \in \mathbb{R}$.

Using set relations together with properties of measurable sets will show the above equivalencies. For example, $a)$ implies $b)$, rewriting the set $\{f \geq a\}$ as

$$\{f \geq a\} = f^{-1}([a, \infty)) = f^{-1}\left(\bigcap_{k=1}^{\infty}(a - 1/k, \infty)\right)$$

and recalling that the inverse image from sets to sets preserves countable intersections (actually arbitrary intersection, unions, and complements) we have:

$$f^{-1}\left(\bigcap_{k=1}^{\infty}(a - 1/k, \infty)\right) = \bigcap_{k=1}^{\infty} f^{-1}\left((a - 1/k, \infty)\right) = \bigcap_{k=1}^{\infty}\{f > a - 1/k\}$$

and the intersection of a sequence of measurable sets is measurable. The following proposition tells us that certain operations performed on measurable functions yield again measurable functions.

**Definition 92.** If $f$ and $g$ are two real-valued functions on $\mathbb{R}$, then we define

$$(f \wedge g)(x) = \min\{f(x), g(x)\}, \qquad (f \vee g)(x) = \max\{f(x), g(x)\}.$$

**Proposition 34.** *Suppose $f$ and $g$ are measurable functions, and suppose the $c \in \mathbb{R}$. Then the functions*

$$cf,\ f + g,\ fg,\ |f|,\ f \wedge g, \quad \text{and} \quad f \vee g \quad \text{are measurable.}$$

*Proof.* We leave it to the reader to prove $cf$ and $fg$ are measurable. For the sum $f + g$ note that if $(f + g)(x) > a$ then there is a rational number $r$ such that $f(x) > r > a - g(x)$, therefore

$$\{f + g > a\} = \bigcup_{r \in \mathbb{Q}} (\{f > r\} \cap \{g > a - r\})$$

which is a countable union of measurable sets. For $|f|$ write

$$\{|f| > a\} = \{f > a\} \cup \{f < -a\}$$

and for $f \wedge g$ and $f \vee g$ we also have:

$$\{f \wedge g > a\} = \{f > a\} \cap \{g > a\},$$

$$\{f \vee g > a\} = \{f > a\} \cup \{g > a\}.$$

□

Now we turn our attention to a collection of measurable functions sharing a common domain $D$ taking values in the extended real line. We are interested in a sequence of measurable functions $f_n : D \to [-\infty, \infty]$.

**Proposition 35.** *Let $(f_n)$ be a sequence of measurable functions, then both* $\sup f_n$ *and* $\inf f_n$ *are measurable.*

*Proof.* Let $a \in \mathbb{R}$; then when we say $\sup_n f_n > a$ we mean $f_n(x) > a$ for some $n$, and conversely

$$\{\sup f_n > a\} = \bigcup_{n=1}^{\infty} \{f_n > a\},$$

which is measurable because each $f_n$ is measurable. Since

$$\inf_n f_n = -\sup_n(-f_n),$$

inf's are measurable since sup's are measurable. □

**Corollary 22.** *Let $(f_n)$ be a sequence of measurable functions. Then both $\limsup_{n\to\infty} f_n$ and $\liminf_{n\to\infty} f_n$ are measurable.*

*Proof.* This directly follows from

$$\limsup_{n\to\infty} f_n = \inf_n \left( \sup_{k\geq n} f_k \right) \quad \text{and} \quad \liminf_{n\to\infty} f_n = \sup_n \left( \inf_{k\geq n} f_k \right). \qquad \square$$

Now we are ready for a major theorem which guarantees that the collection of measurable functions is closed under pointwise limits. This tells us that the class of measurable functions are actually quite large.

**Theorem 111.** *If $(f_n)$ is a sequence of measurable functions and if*

$$f(x) = \lim_{n\to\infty} f_n(x)$$

*exists for all $x \in D$, then the limit function $f$ is measurable.*

*Proof.* Since

$$f(x) = \limsup_{n\to\infty} f_n(x) = \liminf_{n\to\infty} f_n(x),$$

this is a consequence of the above corollary. □

**Remark 61.** Suppose $f$ is a measurable function. Then we divide $f$ into its positive and negative parts as follows:

$$f^+(x) = \sup(f(x), 0) \qquad f^-(x) = \sup(-f(x), 0) = -\inf(f(x), 0).$$

The positive and negative parts of the function are as illustrated in Figure 2.22:

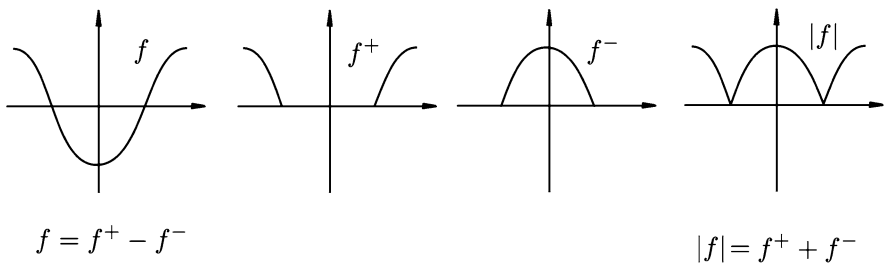



Figure 2.22: Positive and negative parts of a function

Observe that both $f^+$ and $f^-$ are non-negative, measurable and $|f| = f^+ + f^-$; furthermore $f = f^+ - f^-$ provided that this expression makes sense.

**Definition 93** (Almost everywhere)**.** A statement is said to hold *almost everywhere* (a.e.) if it is true except on a set of measure zero.

For example, the characteristic function of the set of rationals defined as:

$$\chi_{\mathbb{Q}}(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \\ 0 & \text{if } x \notin \mathbb{Q} \end{cases}.$$

We have $\chi_{\mathbb{Q}} = 0$ a.e. because it is only nonzero on rationals and the rational numbers being a countable set have measure zero. We shall say

$$f(x) = g(x) \text{ a.e. on a set D} \quad \text{if} \quad m(\{x : f(x) \neq g(x)\}) = 0.$$

It is not hard to show that if $(f_n)$ is a collection of measurable functions and

$$\lim_{n\to\infty} f_n(x) = f(x) \quad \text{a.e.},$$

then $f$ is measurable, i.e., the collection of measurable functions is closed even under pointwise a.e. limits.

**Definition 94.** A *simple function* $\varphi$ is a finite sum

$$\varphi = \sum_{k=1}^{N} a_k \chi_{E_k},$$

where $\chi_{E_k}$ is the characteristic function of $E_k$ defined earlier, each $E_k$ is a measurable set of finite measure and the $a_k$ are constants. The *canonical form* of $\varphi$ is the unique expression as in above where the numbers $a_k$ are distinct and nonzero.

**Example 92.** Let

$$\varphi(x) = \begin{cases} 0 & \text{if } x < 0 \text{ or } x > 2 \\ 1 & \text{if } 0 \leq x \leq 1 \\ 2 & \text{if } 1 < x \leq 2 \end{cases};$$

then we can write $\varphi(x) = \chi_{[0,1]} + 2\chi_{(1,2]} + 0\chi_{(-\infty,0)\cup(2,\infty)}$ which is certainly not in canonical form.

When we define Lebesgue integration we will need the canonical form of $\varphi$. However, for the following theorem on the approximation of measurable functions a general definition of simple functions is enough. The structure of measurable functions provides an insight about their relation to simple functions. The following theorem is about approximating non-negative measurable functions by simple functions.

**Theorem 112** (Approximation by simple functions)**.** *If $f : D \to [0, \infty]$ is a non-negative measurable function, then we can find an increasing sequence of non-negative simple functions $(\varphi_n)$ with*

$$0 \leq \varphi_1 \leq \varphi_2 \leq \cdots \leq f$$

*such that $(\varphi_n)$ converges pointwise to $f$ everywhere on $D$, and that $(\varphi_n)$ converges uniformly to $f$ on any set where $f$ is bounded.*

*Proof.* The general idea of the proof is as follows (readers are advised to fill in the details). Let $f \geq 0$ be a measurable function. For $k \in \mathbb{Z}^+$, $k \neq 0$ divides the interval $[0, k]$ into subintervals

$$\left[\frac{j-1}{2^k}, \frac{j}{2^k}\right] \quad \text{where} \quad j = 1, 2, \ldots, k2^k.$$

Define

$$\varphi_k(x) = \begin{cases} \frac{j-1}{2^k} & \text{if } \frac{j-1}{2^k} \leq f(x) < \frac{j}{2^k} \;\; j = 1, 2, \ldots, k2^k \\ k & \text{if } f(x) \geq k \end{cases}.$$

Observe that each $\varphi_k$ is a simple measurable function defined everywhere in the domain of $f$. Also $\varphi_k \leq \varphi_{k+1}$ because passing from $\varphi_k$ to $\varphi_{k+1}$ each subinterval $\left[\frac{j-1}{2^k}, \frac{j}{2^k}\right]$ is divided by half. The definition of $\varphi_k$ implies that

$$|\varphi_k(x) - f(x)| \leq \frac{1}{2^k} \quad \text{for all} \quad x \in D \quad \text{such that} \quad f(x) \in [0, k]. \qquad \square$$

**Remark 62.** One can prove the above theorem for an arbitrary measurable function $f$, since we have $f = f^+ - f^-$ where both $f^+$ and $f^-$ are non-negative. Applying the above result to each non-negative function we obtain sequences say $(\varphi_k^{'})$ and $(\varphi_k^{''})$ such that

$$\varphi_k^{'} \to f^+ \quad \text{and} \quad \varphi_k^{''} \to f^-.$$

Then $\varphi_k^{'} - \varphi_k^{''}$ is also simple function and $\varphi_k^{'} - \varphi_k^{''} \to f^+ - f^- = f$.

## Littlewood's Three Principles

J. E. Littlewood stated the following principles in his 1944 "Lectures on the Theory of Functions." These principles address the essential issues in measure theory and help us to understand the relationship between the new concepts of measurable sets or measurable functions and convergence of sequence of measurable functions with the older concepts they replaced. These principles are:

a) Every measurable set is "nearly" a finite union of intervals.

b) Every measurable function is "nearly" continuous.

c) Every convergent sequence of measurable functions is "nearly" uniformly convergent.

The first of these principles is about the approximation property of measures. The second item is the well known Lusin's Theorem and the third one is described in a theorem that carries the name Egorov. Of course the key word in the above principles is "nearly," which we explain below when we give the statement of the theorems. First we need a notation. Given two sets $E$ and $F$, the *symmetric difference* between these sets denoted by $E\triangle F$ is defined as

$$E\triangle F = (E \setminus F) \cup (F \setminus E),$$

which consists of points that belong to exactly one of the sets $E$ or $F$.

**Theorem 113** (First Principle)**.** *If $m(E)$ is finite, there exists a finite union $F = \bigcup_{j=1}^{N} Q_j$ of closed intervals such that $m(E\triangle F) \leq \epsilon$.*

**Theorem 114** (Second Principle-Lusin's Theorem)**.** *Suppose $f$ is measurable and finite valued on a set $E$ with the measure of $E$ being finite. Then for every $\epsilon > 0$ there is a closed set $F_\epsilon$ with*

$$F_\epsilon \subset E \quad \text{and} \quad m(E \setminus F_\epsilon) \leq \epsilon$$

*and $f$ restricted to the set $F_\epsilon$, that is $f|_{F_\epsilon}$, is continuous.*

**Theorem 115** (Third Principle-Egorov's Theorem)**.** *Suppose $(f_k)$ is a sequence of measurable functions defined on a measurable set $E$ with $m(E) < \infty$. Assume that $f_k \to f$ a.e. on $E$. Given $\epsilon > 0$, we can find a closed set $F_\epsilon \subset E$ such that*

$$m(E \setminus F_\epsilon) \leq \epsilon \quad \text{and} \quad f_k \to f \quad \text{uniformly on } F_\epsilon.$$

For the proofs of the above three theorems we refer to an excellent text [50], pp. 33–34.

In the following we first return to the Cantor set, after proving that the Cantor set has measure zero, using the construction of the Cantor set, we define Cantor-Lebesgue function. This function is a continuous function $F : [0, 1] \to [0, 1]$ which is increasing with $F(0) = 0$ and $F(1) = 1$. However ,we will see that $F'(x) = 0$ almost everywhere. Hence

$$\int_a^b F'(x)\, dx \neq F(b) - F(a).$$

## Cantor-Lebesgue Function

Recall the Cantor set from Chapter 1. The *Cantor set*, sometimes called the *Cantor ternary set*, is an intriguing subset of $\mathbb{R}$ which will extend our understanding of the nature of subsets of $\mathbb{R}$ and it is a valuable source of counterexamples in Analysis. For example, the Cantor set is the basis of the construction of a function called the *Cantor-Lebesgue function*, sometimes called the "Devil's staircase" or the " Cantor-Scheeffer function," which is a continuous,

non-decreasing function that is not constant, yet has zero derivative at almost every point.

The Cantor set is obtained by successively removing "middle thirds" from the interval $[0, 1]$ as shown in Figure 2.23.

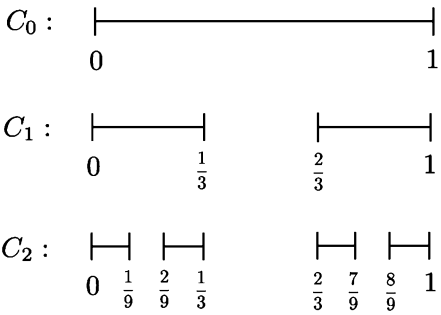



Figure 2.23: Cantor set construction

The Cantor set $C$ is defined as:

$$C = \bigcap_{n=0}^{\infty} C_n.$$

The Cantor set is in some ways large (uncountable) and in other ways small (measure zero).

**Definition 95.** A set $A \subseteq \mathbb{R}$ is said to be *measure zero* if, for $\epsilon > 0$, there exists a countable collection of open intervals $I_n$ such that

$$A \subseteq \bigcup_{n=1}^{\infty} I_n \quad \text{and} \quad \sum_{n=1}^{\infty} |I_n| \leq \epsilon,$$

where by $|I_n|$ we mean the length of the interval $I_n$.

**Proposition 36.** *The Cantor set $C$ has measure zero.*

*Proof.* We start by observing the length of the intervals removed from $[0, 1]$ to form the $C$. To form $C_1$ we removed $I_1 = (\frac{1}{3}, \frac{2}{3})$ and the length of $I_1 = 1/3$. In the second step we removed $I_2 = (\frac{1}{9}, \frac{2}{9}) \cup (\frac{7}{9}, \frac{8}{9})$ and the length of $I_2 = 2\,(1/9)$, and to construct $C_n$ we removed $2^{n-1}$ middle thirds of length $\frac{1}{3^n}$, and so the total length of $[0, 1] \setminus C$ must be:

$$\begin{aligned}
\frac{1}{3} + 2\left(\frac{1}{9}\right) + 4\left(\frac{1}{27}\right) + \cdots + 2^{n-1}\left(\frac{1}{3^n}\right) + \ldots &= \sum_{n=1}^{\infty} \frac{2^{n-1}}{3^n} \\
&= \frac{1}{3} \sum_{n=1}^{\infty} \left(\frac{2}{3}\right)^{n-1} \\
&= \frac{1}{3} \cdot \frac{1}{1 - 2/3} = 1.
\end{aligned}$$

Since $| \bigcup_{n=1}^{\infty} I_n | = 1$, the Cantor set, which is equal to $[0,1] \setminus \bigcup_{n=1}^{\infty} I_n$ has Lebesgue measure zero. Here we used $m(E \setminus F) = m(E) - m(F)$ provided that $m(F) < \infty$. □

Using the Cantor set one can define the following simple construction to describe the so called *Cantor-Lebesgue function*. As above we define the Cantor set

$$C = \bigcap_{k=0}^{\infty} C_k,$$

where each $C_k$ is a disjoint union of $2^k$ intervals, such as

$$C_1 = [0, 1/3] \cup [2/3, 1],$$

$$C_2 = [0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1].$$

**Definition 96.** The Cantor-Lebesgue function $F : [0,1] \to [0,1]$ is a function with $F(0) = 0$ and $F(1) = 1$ defined inductively as follows:

- $F_1(x)$ is a continuous increasing function such that $F_1(0) = 0$, $F_1(1) = 1$, $F_1$ is linear on $C_1$ and $F_1(x) = \frac{1}{2}$ when $x \in [1/3, 2/3]$.
- $F_2(x)$ is a continuous increasing function such that $F_2(0) = 0$, $F_2(1) = 1$, $F_2$ is linear on $C_2$ and

$$F_2(x) = \begin{cases} 1/4 & \text{if } 1/9 \le x \le 2/9, \\ 1/2 & \text{if } 1/3 \le x \le 2/3 \\ 3/4 & \text{if } 7/9 \le x \le 8/9. \end{cases}$$

See Figure 2.24 for $F_1$ and $F_2$:

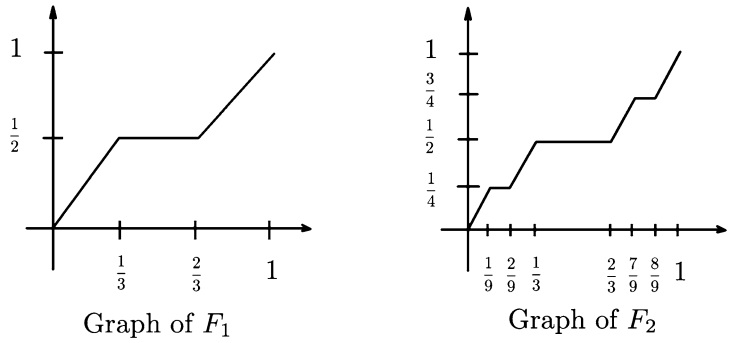



Figure 2.24: Graphs of $F_1$ and $F_2$

This process yields a sequence of continuous increasing functions $(F_n)$ such that $| F_{n+1} - F_n | \le \dfrac{1}{2^{n+1}}$ and $F_n \to F$ uniformly; thus the limit function $F$ is also continuous and called the *Cantor-Lebesgue function* (Figure 2.25).

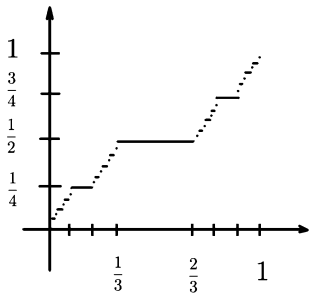



Figure 2.25: The Cantor-Lebesgue function

Note that this function can also be defined by converting base 3 representation into base 2. The following are some well-known consequences:

a) $F$ is constant on each interval of the complement of the Cantor set. Since $m(C) = 0$, we find that $F'(x) = 0$ a.e.. This is particularly interesting in the context of the Fundamental Theorem of Calculus. We see that the Cantor-Lebesgue function does not satisfy

$$F(b) - F(a) = \int_a^b F'(x)\,dx,$$

since $F(1) - F(0) = 1 \neq 0$.

b) A continuous image of a measurable set can be a non-measurable set. One can see this using the Cantor-Lebesgue function. To understand this, first observe that $F(C) = [0, 1]$ and recall that not every subset of $\mathbb{R}$ is measurable (see the section on the Banach–Tarski paradox). If $E$ is a subset of $[0, 1]$ that is not Lebesgue measurable, and if we let

$$B = C \cap F^{-1}(E),$$

then $m(B) = 0$ because $B \subset C$ and $m(C) = 0$. But every subset of $\mathbb{R}$ with Lebesgue measure zero is measurable, thus $B$ is measurable and $F(B) = E$.

## Lebesgue Integration

For the Riemann integral, we divide the area under the graph of $f$ into vertical rectangles to find upper and lower sums. How about dividing this area into horizontal rectangles? This leads to a more complicated mathematics. Recalling Remark 30 in Section 1.8, suppose $f : [a, b] \to \mathbb{R}$ to be a non-negative bounded function defined on $[a, b]$. Let $R = \{y_0 < y_1 < \cdots < y_n\}$ be a partition of the range of $f$ as shown in Figure 2.26.

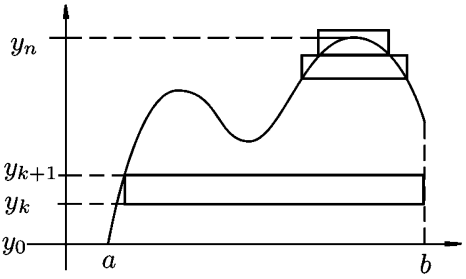



Figure 2.26: Dividing by horizontal rectangles

Suppose we want to find

$$\sum_{k=1}^{n}(y_{k+1}-y_k)\ (\text{length of the interval}\,\{x : f(x)\geq y_k\}).$$

However, the set $\{x : f(x) \geq y_k\}$ might be a complicated set and one asks how can one find its length? To answer this question one has to develop the idea of measure (or length) of a set. That is why we introduced the concept of measure first. Among many notions of integral the most prominent one is the Lebesgue integral. To develop these ideas precisely requires a text in itself (see [8, 22, 50]). However, here we can answer some basic questions. Lebesgue integral defined by a step by step process, and in each step one is able to integrate larger families of functions. We define Lebesgue integral of simple functions first, then we move on to bounded functions supported on a set of finite measure, then the Lebesgue integral of non-negative functions is introduced before we go to general case which is the class of integrable functions. We start with the building blocks for the theory of integration, namely the characteristic function $\chi_E$ of $E$ (sometimes called the indicator function of $E$) which is

$$\chi_E(x) = \begin{cases} 1 & \text{if } x \in E \\ 0 & \text{if } x \notin E \end{cases}.$$

It is easy to see that the height of $\chi_E$ is one; therefore, the area under the function $\chi_E$ must be the length of $E$, or more precisely $m(E)$. We denote this by $\int_E \chi_E\,dx$. This implies that the area under $a\chi_E$ is then equal to

$$a\,m(E) = \int_E a\,\chi_E\,dx.$$

Now consider two disjoint sets $E$ and $F$, with the same arguments the area under the function $a\,\chi_E + b\,\chi_F$ must be equal to $a\,m(E) + b\,m(F)$, and we write this as (Figure 2.27):

$$\int_{E\cup F}(a\,\chi_E + b\,\chi_F)\,dx = a\,m(E) + b\,m(F).$$

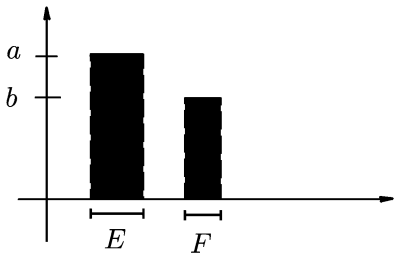



Figure 2.27: Total area $= a\,m(E) + b\,m(F)$

## The Lebesgue Integral of Simple Functions

Suppose $\varphi$ is a simple function in the canonical form $\sum_{k=1}^{N} a_k \chi_{E_k}$ where the $E_k$ are measurable sets of finite measure and the $a_k$'s are distinct and nonzero, (canonical form assures that decomposition $\varphi$ into sum is unique). In this case we define the *Lebesgue integral* of $\varphi$ by

$$\int_{\mathbb{R}} \varphi\, dx = \sum_{k=1}^{N} a_k m(E_k).$$

If $F$ is a measurable subset of $\mathbb{R}$ with finite measure, we can define the Lebesgue integral of $\varphi$ over $F$ by

$$\int_F \varphi\, dx = \int \varphi(x)\chi_F(x)\, dx.$$

We simply use $\int \varphi\, dx$ for the integral of $\varphi$ over $\mathbb{R}$. If clarity requires the variable to be emphasized, then the notation $\int \varphi(x)\, dx$ will be used. Note that if we need to use representations of $\varphi$ which are not canonical then we need to prove $\int_{\mathbb{R}} \varphi\, dx = \sum_{k=1}^{N} a_k m(E_k)$ is independent of any representation of $\varphi$. For the proof of this claim see [50], p. 51.
The integral of simple functions defined above satisfies the following properties:

a) If $\varphi$ and $\psi$ are simple and $a, b \in \mathbb{R}$, then

$$\int (a\varphi + b\psi)\, dx = a \int \varphi\, dx + b \int \psi\, dx \quad \text{(Linearity)}.$$

b) If $E$ and $F$ are disjoint subsets of $\mathbb{R}$ with finite measure, then

$$\int_{E\cup F} \varphi\, dx = \int_E \varphi\, dx + \int_F \varphi\, dx \quad \text{(Additivity)}.$$

c) If $\varphi \leq \psi$ are simple, then

$$\int \varphi \, dx \leq \int \psi \, dx \quad \text{(Monotonicity)}.$$

d) If $\varphi$ is a simple function, then so is $|\varphi|$, and

$$\left| \int \varphi \right| dx \leq \int |\varphi| \, dx \quad \text{(Triangle Inequality)}.$$

## The Lebesgue Integral of a Bounded Function over a Set of Finite Measure

**Definition 97.** Let $f$ be a measurable function, then the *support* of $f$ is the set of points $x$ for which $f(x) \neq 0$. In other words, $\text{supp}(f) = \{x : \ f(x) \neq 0\}$. Note that if $f$ is measurable so is the set $\text{supp}(f)$.

Recalling the theorem of approximating measurable functions, Theorem 112, if a function $f$ is supported on a set $E$ and bounded say by a constant $M$, then

a) There exists a sequence of simple functions $(\varphi_n)$ such that $\varphi_n(x) \to f(x)$ for all $x$.

b) Each $\varphi_n$ is bounded on $E$ by $M$.

c) Each $\varphi_n$ is supported on $E$.

The importance of these cases stems from Egorov's result (see Theorem 115) which states that the convergence of a sequence of measurable functions is "nearly" uniform. Now we give the definition of integration for bounded functions that are supported on a set of finite measure. For such a function we define its Lebesque integral by:

$$\int f(x) \, dx = \lim_{n \to \infty} \int \varphi_n(x) \, dx,$$

where $\{\varphi_n\}$ is any sequence of simple functions satisfying the conditions $a), b)$, and $c)$ stated above. However we do not know if this limit exists. Egorov's theorem can be applied here. Since the measure of $E$ is finite, given $\epsilon > 0$ there exists a closed measurable subset $E_\epsilon$ of $E$ such that $m(E \setminus E_\epsilon) \leq \epsilon$ and $\varphi_n \to f$ uniformly on $E_\epsilon$. Therefore by setting $I_n = \int \varphi_n \, dx$, we can estimate $|I_n - I_m|$ as

$$|I_n - I_m| \leq \int_{E_\epsilon} |\varphi_n - \varphi_m| \, dx + \int_{E \setminus E_\epsilon} |\varphi_n - \varphi_m| \, dx \leq \int_{E_\epsilon} |\varphi_n - \varphi_m| \, dx + 2M\epsilon.$$

By the uniform convergence, we have, for all $x \in E_\epsilon$ and for all sufficiently large $n$ and $m$, the estimate $|\varphi_n - \varphi_m| < \epsilon$, so we conclude that

$$|I_n - I_m| \leq m(E)\epsilon + 2M\epsilon$$

for all large $n$ and $m$. Since $m(E) < \infty$ and $\epsilon$ is arbitrary, this shows that $\{I_n\}$ is a Cauchy sequence and hence converges. Therefore $\lim_{n\to\infty} \int \varphi_n(x)\,dx$ exists. Furthermore, if $f = 0$ a.e., then repeating the argument above one can show $|I_n| \leq (m(E) + 2M)\epsilon$ which yields $\lim_{n\to\infty} \int \varphi_n(x)\,dx = 0$. It is not difficult to show that $\int f\,dx$ is independent of the limiting sequence $\varphi_n$ used in order for the integral to be well defined. (Why?) Furthermore if $f$ is a simple function itself, then $\int f\,dx$ as defined above coincides with the integral of simple functions given before. Thus one can extend the basic properties of integration defined for simple functions, to bounded functions supported on a set of finite measure.

**Proposition 37.** *Suppose $f$ and $g$ are bounded functions supported on a set of finite measure. Then the following properties hold.*

*a) If $a, b \in \mathbb{R}$, then*

$$\int (af + bg)\,dx = a \int f\,dx + b \int g\,dx \quad \textit{(Linearity).}$$

*b) If $E$ and $F$ are disjoint subsets of $\mathbb{R}$ with finite measure, then*

$$\int_{E\cup F} f\,dx = \int_E f\,dx + \int_F g\,dx \quad \textit{(Additivity).}$$

*c) If $f \leq g$, then*

$$\int f\,dx \leq \int g\,dx \quad \textit{(Monotonicity).}$$

*d) If $f$ is bounded then $|f|$ is also bounded, supported on a set of finite measure, and*

$$\left|\int f\,dx\right| \leq \int |f|\,dx \quad \textit{(Triangle Inequality).}$$

All of the properties follow directly from using approximation of measurable functions by simple functions, and the properties of the integral of simple functions were given previously.

For bounded functions supported on a set of finite measure we have a crucial convergence theorem. This theorem is about the interchange of an integral and a limit. Because its conclusion is simply:

$$\lim_{n\to\infty} \int f_n\,dx = \int \lim_{n\to\infty} f_n\,dx.$$

**Theorem 116** (Bounded Convergence Theorem (BCT))**.** *Suppose $(f_n)$ is a sequence of measurable functions such that $|\ f_n(x)\ | \leq M$ for all $n$ and $(f_n)$ supported on a set $E$ of finite measure. Suppose also $f_n(x) \to f(x)$ a.e. as $n \to \infty$. Then the limit function $f$ is measurable, bounded, supported on $E$ for a.e. $x$, and*

$$\int |f_n - f|\,dx \to 0 \quad \textit{as} \quad n \to \infty.$$

*Proof.* By Egorov's Theorem 115, we know there is measurable subset $E_\epsilon$ of $E$ such that $m(E \setminus E_\epsilon) \leq \epsilon$ and $f_n \to f$ uniformly on $E_\epsilon$. Thus for all sufficiently large $n$ we must have

$$| f_n(x) - f(x) | \leq \epsilon \quad \text{for all} \quad x \in E_\epsilon.$$

Using the properties of the integral,

$$\int | f_n - f | \, dx \leq \int_{E_\epsilon} | f_n - f | \, dx + \int_{E \setminus E_\epsilon} | f_n - f | \, dx \leq \epsilon \, m(E) + 2M m(E \setminus E_\epsilon)$$

for all $n$ large enough. Since $\epsilon$ is arbitrary, the proof is completed. □

Observe from $\int | f_n - f | \, dx \to 0$ as $n \to \infty$ we conclude $\int f_n \, dx \to \int f \, dx$ as $n \to \infty$. This theorem will be developed further in the section on "convergence theorems." Another useful observation at this point is the fact that if $f \geq 0$, bounded and a supported on a set of finite measure $E$ and $\int f \, dx = 0$ then $f = 0$ almost everywhere. For the proof see part $f$) of Proposition 38 below.

**Remark 63.** If we assume $f$ is Riemann integrable on a closed interval $[a, b]$, then $f$ is measurable, and the standard Riemann integral, is equal to its Lebesgue integral. In other words

$$(R) \int_{[a,b]} f \, dx = (L) \int_{[a,b]} f \, dx,$$

where the integral on the left-hand side is the Riemann Integral, and the one on the right-hand side is the Lebesgue integral. (See the exercises at the end of this section). This result combined with bounded convergence theorem shows why one can integrate much larger class of functions in the sense of Lebesgue.

## The Lebesgue Integral of Non-Negative Functions

We next define the integral of non-negative measurable but not necessarily bounded functions. We will allow these functions to be extended-valued, i.e., these functions may take the value $+\infty$ on a measurable set. It is useful to remember the convention that one defines the supremum of a set of positive numbers to be $+\infty$ if the set is unbounded.

**Definition 98.** Let $f : \mathbb{R} \to [0, \infty]$ be measurable, we define the *Lebesgue integral* of $f$ over $\mathbb{R}$ by

$$\int f(x) \, dx = \sup_g \int g(x) \, dx,$$

where the supremum is taken over all measurable functions $g$ such that $g$ is bounded and supported on a set of finite measure and $0 \leq g \leq f$.

In the above definition of the integral, there are two possible cases: the supremum is either finite or infinite, if $\int f\,dx < \infty$, then we say $f$ is *Lebesgue integrable* on $\mathbb{R}$. It follows that if $E$ is a measurable subset of $\mathbb{R}$, and $f \geq 0$ then $f\chi_E$ is also positive and we define

$$\int_E f(x)\,dx = \int f(x)\chi_E(x)\,dx.$$

Note that we are not excluding the possibility that $\int f(x)\,dx = \infty$. In particular, if $E$ is measurable and $m(E) < \infty$ then as usual $m(E) = \int \chi_E\,dx$. However when $m(E) = \infty$, we have

$$\int \chi_E\,dx \geq \sup_n \int \chi_{E\cap[-n,n]}\,dx = \sup_n m(E \cap [-n,n]) = m(E) = \infty.$$

**Proposition 38.** *The integral of non-negative measurable functions satisfies the following properties:*

*a) Linearity. If $f, g \geq 0$, and $a, b$ are positive real numbers, then*

$$\int (af + bg)\,dx = a\int f\,dx + b\int g\,dx.$$

*b) Additivity. If $E$ and $F$ are disjoint subsets of $\mathbb{R}$, and $f \geq 0$, then*

$$\int_{E\cup F} f\,dx = \int_E f\,dx + \int_F f\,dx.$$

*c) Monotonicity. If $0 \leq f \leq g$ then*

$$\int f\,dx \leq \int g\,dx.$$

*d) If $g$ is integrable and $0 \leq f \leq g$, then $f$ is integrable as well.*

*e) If $f$ is integrable, then $f(x) < \infty$ for almost every $x$.*

*f) If $\int f\,dx = 0$, then $f(x) = 0$ for almost every $x$.*

*Proof.* We only give proofs of parts e) and f). Parts b), c), and d) follow from the definitions; part a) is not immediate but we leave to the reader.

To prove part e), set up the sets $E_k = \{x : f(x) \geq k\}$ and $E_\infty = \{x : f(x) = \infty\}$. From the set $E_k$ we can write

$$k\,m(E_k) \leq \int \chi_{E_k} f\,dx \leq \int f\,dx$$

and hence $m(E_k) \to 0$ as $k \to \infty$. Since $E_k$ decreases to $E_\infty$, using the continuity property of measures we obtain $m(E_\infty) = \lim_{k\to\infty} m(E_k) = 0$.

To prove f), if $f \geq 0$, supported on a set of finite measure $E$ and $\int f\,dx = 0$ then we want to show $f = 0$ almost everywhere. For each integer $k \geq 1$ define $E_k = \{x \in E : \ f(x) \geq 1/k\}$; then from the inequality $k^{-1}\chi_{E_k} \leq f$ and the monotonicity of the integral we see that

$$m(E_k) \leq \frac{1}{k}\int f\,dx.$$

Thus $m(E_k) = 0$ for all $k$, and since

$$\{x : \ f(x) > 0\} = \bigcup_{k=1}^{\infty} E_k,$$

we obtain the desired result that $f = 0$ almost everywhere. □

## General Lebesgue Integral

Suppose $f$ is a measurable function. Then we divide $f$ into its positive and negative parts as follows:

$$f^+(x) = \sup(f(x), 0) \qquad f^-(x) = \sup(-f(x), 0) = -\inf(f(x), 0).$$

**Definition 99.** A measurable function $f$ is *integrable* if the non-negative measurable function $|f|$ is integrable in the sense of previous section.
That is, if $\int \mid f \mid\, dx < \infty$.

Note that the functions $f^{\pm}$ satisfy $0 \leq f^{\pm} \leq \mid f \mid$, so $f$ is integrable if and only if $\int f^+\,dx < \infty$ and $\int f^-\,dx < \infty$.

**Definition 100.** If $f$ is integrable then its integral is

$$\int f\,dx = \int f^+\,dx - \int f^-\,dx.$$

This is consistent with the earlier definition, since $f \geq 0$ implies $f^+ = f$ and $f^- = 0$.

**Theorem 117** (Basic Properties of the Integral)**.** *The integral of Lebesgue integrable functions is linear, additive, monotonic and satisfies the triangle inequality.*

*Proof.* These properties require proof, but we only prove additivity, since $(f + g)^{\pm} \neq f^{\pm} + g^{\pm}$. However if $f$ and $g$ are integrable, so is $f + g$ since we have the triangle inequality $\mid f + g \mid \leq \mid f \mid + \mid g \mid$. Next observe that

$$(f + g)^+ - (f + g)^- = f + g = (f^+ - f^-) + (g^+ - g^-).$$

Therefore

$$(f + g)^+ + f^- + g^- = (f + g)^- + f^+ + g^+.$$

All of the functions in the above identity are non-negative so we can apply additivity to obtain

$$\int (f+g)^+ \, dx + \int f^- \, dx + \int g^- \, dx = \int (f+g)^- \, dx + \int f^+ \, dx + \int g^+ \, dx,$$

and so

$$\int (f+g)^+ \, dx - \int (f+g)^- \, dx = \int (f+g) \, dx = \int (f^+ - f^-) \, dx + \int (g^+ - g^-) \, dx$$

yielding

$$\int (f+g) \, dx = \int f \, dx + \int g \, dx. \qquad \square$$

**Remark 64.** If $f$ is integrable on $\mathbb{R}$, $E$ is a measurable set and $f$ is defined on $E$, the integral of the restriction over $E$ is defined as $\int f\chi_E \, dx$, and $f\chi_E = 0$ on $E^c$, the complement of $E$. When defining the product $f\chi_E$ we use $0(\pm\infty) = 0$.

**Remark 65** (Tchebychev Inequality)**.** Suppose $f \geq 0$ and integrable. If $\lambda > 0$ and $E_\lambda = \{x : \ f(x) > \lambda\}$, then one can show

$$m(E_\lambda) \leq \frac{1}{\lambda} \int f \, dx.$$

The importance of this inequality is that it estimates the "size" of $f$ in terms of the integral of $f$.

## The Space $L^1$ of Integrable Functions

The set of integrable functions form a vector space, so it is appropriate to ask how one can place a norm on it so it becomes a complete normed space.

**Definition 101.** For any integrable function $f$ on $\mathbb{R}$, we define the $L^1$- *norm* of $f$ as:

$$|| \, f \, || = || \, f \, ||_{L^1} = \int_{\mathbb{R}} |f(x)| \, dx.$$

Note that $|| \, f \, || = 0$ if and only if $f = 0$ a.e.. Furthermore, we do not distinguish between two functions that agree almost everywhere; with these facts in mind we define $L^1(\mathbb{R})$ as the set of all equivalence classes of integrable functions (two functions $f$ and $g$ are equivalent if $f = g$ a.e.). We list the elementary facts about the norm $|| \, f \, ||_{L^1}$ in the following proposition.

**Proposition 39.** *Let $f, g \in L^1(\mathbb{R})$, then*

*a)* $|| \, f \, ||_{L^1} = 0$ *if and only if* $f = 0$ *a.e..*

*b)* $|| \, kf \, ||_{L^1} = |k| || \, f \, ||_{L^1}$ *for all scalars* $k$*.*

*c)* $|| \, f + g \, ||_{L^1} \leq || \, f \, ||_{L^1} + || \, g \, ||_{L^1}$*.*

This norm induces a metric $d(x, y) = || f - g ||_{L^1(\mathbb{R})}$ on $L^1(\mathbb{R})$. There are two properties of $L^1(\mathbb{R})$ we now mention without proof (see [50], p. 70–71):

- The vector space $L^1(\mathbb{R})$ is complete in its metric (called the *Riesz-Fisher Theorem*).
- The families of simple functions, step functions, and continuous functions with compact support are dense in $L^1(\mathbb{R})$.

## Convergence Theorems

In this section we see three fundamental convergence theorems in Lebesgue integration theory. Namely the dominated convergence theorem (DCT), monotone convergence theorem (MCT), and Fatou's lemma. In all of these theorems we have a sequence of functions $(f_n)$ such that $f_n \to f$ a.e. Then Fatou's lemma, MCT, and DCT all state under suitable conditions that we can assert something about $\int f\,dx$ in terms $\int f_n\,dx$. Fatou's lemma has the weakest hypothesis; we need to only have $f_n$ bounded below by 0. Consequently its conclusion is weaker; we can only conclude

$$\int f\,dx \leq \liminf \int f_n\,dx.$$

In DCT, we assume $f_n$ is bounded above and below by an integrable function $g$, i.e., $\mid f_n(x) \mid < g(x)$, then we conclude

$$\int f\,dx = \lim \int f_n\,dx.$$

In MCT we require $f_n$ to be bounded below by zero and bounded above by the limit function $f$. In other words we assume $0 \leq f_1 \leq f_2 \leq \cdots \leq f$ and conclude

$$\int f\,dx = \lim \int f_n\,dx.$$

Clearly, if the limit function $f$ is integrable then the MCT becomes a special case of DCT. The advantage of Fatou's lemma and MCT is that they are applicable even if $f$ is not integrable and are often a good way to show that $f$ is integrable.

**Lemma 13** (Fatou)**.** *Suppose $(f_n)$ is a sequence of measurable functions with $f_n \geq 0$. If $f_n \to f$ a.e., then*

$$\int f\,dx \leq \liminf_{n\to\infty} \int f_n\,dx.$$

Before we prove Fatou's lemma we observe the following points:

- The conclusion of Fatou's lemma can be written as

$$\int \liminf_{n\to\infty} f_n\,dx \leq \liminf_{n\to\infty} \int f_n\,dx.$$

- In case $\lim_{n\to\infty}\int f_n$ exists then Fatou's lemma assures that

$$\int \lim_{n\to\infty} f_n\,dx \le \lim_{n\to\infty}\int f_n\,dx.$$

- $f_n(x)\to f(x)$ a.e. means $\lim_{n\to\infty} f_n(x) = f(x)$ for every $x\in E$ and $m(E^c)=0$. Therefore

$$\begin{aligned}\int f\,dx &= \int_E f\,dx + \int_{E^c} f\,dx = \int_E f\,dx\\ &= \int f\chi_E\,dx \le \liminf_{n\to\infty}\int f_n\chi_E\,dx = \liminf_{n\to\infty}\int_E f_n\,dx.\end{aligned}$$

*Proof.* Suppose $g$ is a bounded function and supported on a set $E$ of finite measure with $0\le g\le f$. Set

$$g_n(x) = \min(g(x), f_n(x)).$$

Then

- $g_n$ is measurable and supported on $E$.
- $g_n(x)\le f_n(x)$ for all $x$ and thus $\int g_n\,dx \le \int f_n\,dx$.
- $g_n(x)\to g(x)$ a.e..

We can apply Theorem 116 (BCT) to $g_n$ to conclude

$$\lim_{n\to\infty}\int g_n\,dx = \int \lim_{n\to\infty} g_n\,dx.$$

Therefore,

$$\int g\,dx = \lim_{n\to\infty}\int g_n\,dx \le \lim_{n\to\infty}\int f_n\,dx$$

and $\int g\,dx \le \liminf_{n\to\infty}\int f_n\,dx$. Taking the supremum over all $g$ with $g\le f$ gives the claimed inequality. □

The following are some consequences of Fatou's lemma.

**Corollary 23.** *Suppose $f\ge 0$ is a non-negative measurable function, and $\{f_n\}$ a sequence of non-negative measurable functions with $f_n(x)\le f(x)$ and $f_n(x)\to f(x)$ a.e.. Then*

$$\lim_{n\to\infty}\int f_n\,dx = \int f\,dx.$$

*Proof.* Since $f_n(x) \le f(x)$ a.e., we have $\int f_n(x)\,dx \le \int f(x)\,dx$ for all $n$; hence

$$\limsup_{n\to\infty} \int f_n\,dx \le \int f\,dx.$$

By Fatou's lemma we also have

$$\int f\,dx \le \liminf_{n\to\infty} \int f_n\,dx,$$

and combining these two inequalities proves the claim. □

**Corollary 24** (Monotone Convergence Theorem)**.** *Suppose $(f_n)$ is a sequence of non-negative measurable functions that satisfies*

$$\lim_{n\to\infty} f_n(x) = f(x) \text{ a.e } x. \quad \text{and} \quad f_n(x) \le f_{n+1}(x) \text{ a.e } x \text{ all } n \ge 1.$$

*Then*

$$\lim_{n\to\infty} \int f_n\,dx = \int f\,dx.$$

*Proof.* Suppose $(f_n)$ is a sequence of measurable functions that satisfies the given conditions. Since $(f_n)$ is increasing, $\lim_{n\to\infty} f_n(x) = \sup_n f_n(x)$ and

$$0 \le \int f_n\,dx \le \int f_{n+1}\,dx \le \int f\,dx.$$

Therefore $\limsup_{n\to\infty} \int f_n\,dx \le \int f\,dx$, and here again apply Fatou's lemma to get the desired conclusion. □

**Remark 66.** MCT implies the linearity property of the integral of non-negative measurable functions. Because if $f$ and $g$ are two such functions we can choose two sequences of measurable functions $\varphi_n$ and $\psi_n$, both supported on a set of finite measure and $\lim_{n\to\infty} \varphi_n = f$ a.e. and $\lim_{n\to\infty} \psi_n = g$ a.e.. Then $\varphi_n + \psi_n$ increases to $f + g$ and by the MCT

$$\int (f+g)\,dx = \lim_{n\to\infty} \int (\varphi_n + \psi_n)\,dx.$$

Now since both $\varphi_n$ and $\psi_n$ are bounded we can apply BCT to interchange limit and integral in the above equality to conclude

$$\int (f+g)\,dx = \int f\,dx + \int g\,dx.$$

**Corollary 25** (Beppo-Levi Theorem)**.** *Consider an infinite series of the form $\sum_{k=1}^{\infty} a_k(x)$ where $a_k(x) \ge 0$ and is measurable for every $k \ge 1$. Then*

$$\int \sum_{k=1}^{\infty} a_k(x)\,dx = \sum_{k=1}^{\infty} \int a_k(x)\,dx.$$

*Proof.* Let $f_n(x) = \sum_{k=1}^{n} a_k(x)$ and $f(x) = \sum_{k=1}^{\infty} a_k(x)$; then the functions $f_n$ are measurable, $f_n(x) \leq f_{n+1}(x)$ and $f_n(x) \to f(x)$ as $n \to \infty$ so we can apply MCT. □

**Example 93.** Suppose we want to show that

$$\lim_{n\to\infty} \int_0^n \left(1 + \frac{x}{n}\right)^n \, e^{-2x}\, dx = 1.$$

First note that $e^{-x}$ is Lebesgue integrable over $[0, \infty)$, because

$$\int_0^\infty e^{-x}\, dx = \lim_{t\to\infty} \int_0^t e^{-x}\, dx = \lim_{t\to\infty} (1 - e^{-t}) = 1.$$

Setting

$$g_n(x) = \left(1 + \frac{x}{n}\right)^n \, e^{-2x} \, \chi_{[0,n]}(x)$$

and recalling from elementary calculus that $\left(1 + \frac{x}{n}\right)^n \nearrow e^x$ for each $x \geq 0$ we see that

$$g_n(x) \nearrow e^{-x} \qquad \text{holds for each} \quad x \geq 0.$$

We can apply MCT and interchange the limit and integral to obtain

$$\lim_{n\to\infty} \int_0^n \left(1 + \frac{x}{n}\right)^n \, e^{-2x}\, dx = \lim_{n\to\infty} \int g_n(x)\, dx = \int \lim_{n\to\infty} g_n(x)\, dx = \int_0^\infty e^{-x}\, dx = 1.$$

Next we prove the most important theorem of the theory of Lebesgue integration, the dominated convergence theorem. For its proof we need the following lemma:

**Lemma 14.** *Suppose $f$ is an integrable function on $\mathbb{R}$. Then for every $\epsilon > 0$ there exists a set $E$ of finite measure such that*

$$\int_{E^c} |\, f \,|\, dx < \epsilon.$$

*Proof.* Without loss of generality assume $f \geq 0$ (if not replace $f$ by $|f|$). Let us take $E = B(0; N) = B_N$ as the ball of radius $N$ centered at the origin, and the truncation of $f$ denoted by $f_N$ defined as

$$f_N = f\, \chi_{B_N}.$$

Then the $f_N \geq 0$ is measurable, increasing and $f_N \to f$ as $N \to \infty$. Apply MCT to $f_N$ to get

$$\lim_{N\to\infty} \int f_N\, dx$$
$$= \int f\, dx \quad \text{and in particular for some large } N, \quad 0 \leq \int f\, dx - \int f\, \chi_{B_N}\, dx < \epsilon$$

or

$$0 \leq \int (1 - \chi_{B_N}) f \, dx = \int \chi_{B_N^c} f \, dx = \int_{B_N^c} f \, dx < \epsilon. \qquad \square$$

**Theorem 118** (Dominated Convergence Theorem)**.** *Suppose $(f_n)$ is a sequence of measurable functions such that $f_n(x) \to f(x)$ a.e. as $n \to \infty$. If*

$$|f_n(x)| \leq g(x) \text{ where } g \text{ is integrable,}$$

*then*

$$\lim_{n\to\infty} \int f_n \, dx = \int f \, dx.$$

*Proof.* For each $N$ define $E_N = \{x : |x| \leq N, \ g(x) \leq N\}$. Given $\epsilon > 0$, by the Lemma 14 and $g$ being integrable there exists some $N$ such that $\int_{E_N^c} g \, dx < \epsilon$. Moreover, the function $f \chi_{E_N} \leq N$ and is supported on a set of finite measure. Thus we can apply BCT to say

$$\int_{E_N} | f_n - f | \, dx < \epsilon \quad \text{for all large } n.$$

Hence

$$\begin{aligned} \int |f_n - f| \, dx &= \int_{E_N} |f_n - f| \, dx + \int_{E_N^c} |f_n - f| \, dx \\ &\leq \int_{E_N} |f_n - f| \, dx + 2 \int_{E_N^c} g \, dx \\ &\leq \epsilon + 2\epsilon. \end{aligned}$$

$\square$

**Remark 67.** It is possible to weaken the hypothesis of DCT to $|f_n(x)| \leq g(x)$ a.e. and that $f_n \to f$ a.e.. See also Exercise 10 below for another version.

## Exercises

1. Prove that if $a, b \in \mathbb{R}$ and $a < b$, then $|(a, b)| = |[a, b)| = |(a, b]| = b - a$.

2. Prove that if $f$ and $g$ are measurable, then so is $cf$ and $fg$ where $c$ is real.

3. Let $\phi$ be continuous on $\mathbb{R}$ and let $f$ be finite a.e. in $E \subset \mathbb{R}^d$, so that in particular $\phi \circ f$ is defined a.e. in $E$.

   (a) Show that $\phi(f)$ is measurable if $f$ is.

   (b) Show that $|f|, |f|^p$ for $p > 0$, and $e^{cf}$ are measurable if $f$ is.

   (c) Give an example of a function $f$ which is not measurable but $|f|$ is measurable.

4. Let $A$ be a dense subset of $\mathbb{R}$. Show that $f$ is measurable if $\{x : \ f(x) > a\}$ is a measurable set for all $a \in A$.

5. Show that

a) The sum and product of two simple functions are simple.

b) $\chi_{A\cap B} = \chi_A \cdot \chi_B$.

c) $\chi_{A\cup B} = \chi_A + \chi_B - \chi_{A\cap B}$.

d) $\chi_{A^c} = 1 - \chi_A$.

6. (Tchebychev Inequality). Suppose $f \geq 0$ and integrable. If $\lambda > 0$ and $E_\lambda = \{x : \ f(x) > \lambda\}$, then show that

$$m(E_\lambda) \leq \frac{1}{\lambda} \int f \, dx.$$

7.

a) Give an example of a function $f : \mathbb{R} \to \mathbb{R}$ which is integrable, but $f^2$ is not.

b) Show that a measurable function $f$ is integrable if and only if $|f|$ is integrable. Give an example of a nonintegrable function whose absolute value is integrable.

8. Suppose $f$ is Riemann integrable on the closed interval $[a, b]$. Then show that $f$ is measurable and

$$(R) \int_{[a,b]} f \, dx = (L) \int_{[a,b]} f \, dx,$$

where the integral on the left-hand side is the Riemann Integral, and the one on the right-hand side is the Lebesgue integral.

9. Let $\mu$ be the counting measure on $\mathbb{N}$. Interpret Fatou's lemma and the monotone and dominated convergence theorems as statements about infinite series.

10. If $f_n, g_n, f, g$ are all integrable functions with $f_n \to f$ and $g_n \to g$ a.e., $|f_n| \leq g_n$, and $\int g_n \, dx \to \int g \, dx$, then show that $\int f_n \, dx \to \int f \, dx$.

11. Suppose $f_n, f \in L^1$ and $f_n \to f$ a.e.. Then show that $\int |f_n - f|\, dx \to 0$ if and only if $\int |f_n|\, dx \to \int |f|\, dx$.

12. Suppose $f \geq 0$; let

$$\mu(E) = \int_E f\, dm$$

for a measurable set $E$.

a) Show that $\mu$ is a measure.

b) Show that for any $g \geq 0$

$$\int g\, d\mu = \int fg\, dm.$$

Hint: First suppose $g$ is simple.

13. Show that $f(x) = \dfrac{\ln x}{x^2}$ is Lebesgue integrable over $[1, \infty)$ and that $\int f dx$=1.

14. Establish the Riemann–Lebesgue Theorem: if $f$ is an integrable function on $(-\infty, \infty)$ then

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} f(x) \cos(nx)\, dx = 0.$$

Hint: The theorem is easy if $f$ is a simple function, then use Theorem 2.4 in [50], p. 71.

15. Let $f$ be integrable over $(-\infty, \infty)$, then show that

a)

$$\int f(x)\, dx = \int f(x + t)\, dx.$$

b) Let $g$ be a bounded measurable function. Then show that

$$\lim_{t\to 0} \int_{-\infty}^{\infty} |g(x)[f(x) - f(x + t)]|\, dx = 0.$$

16.

a) Let $\{f_n\}$ be a sequence of real-valued measurable functions. If $\{f_n\}$ converges to $f$ in measure, show that $\{f_n\}$ is a Cauchy sequence in measure.

b) Show that if a sequence $\{f_n\}$ of integrable functions converge to $f$ in $L^1$, then $\{f_n\}$ converges $f$ in measure. Is the converse true?

Note: We say $\{f_n\}$ is *Cauchy in measure* if for every $\epsilon > 0$,

$$m(\{x : \ |f_n(x) - f_m(x)| > \epsilon\}) \to 0 \quad \text{as } m, n \to \infty$$

and we say $\{f_n\}$ *converges in measure* to a measurable function $f$ if for every $\epsilon > 0$

$$m(\{x : \ |f_n(x) - f(x)| > \epsilon\}) \to 0 \quad \text{as } n \to \infty.$$

17. Compute the following limits and justify the calculations:

a) $\lim_{n\to\infty} \int_0^\infty [1 + (x/n)]^{-n} \sin(x/n) dx.$

b) $\lim_{n\to\infty} \int_0^\infty n \sin(x/n)[x(1 + x^2)]^{-1} dx.$

c) $\lim_{n\to\infty} \int_a^\infty n(1 + n^2x^2)^{-1} dx.$

18. Show that $\int_0^\infty x^{2n} e^{-x^2} dx = \dfrac{(2n)!}{2^{2n} n!} \cdot \dfrac{\sqrt{\pi}}{2}$ holds true for $n = 0, 1, 2, \ldots$

Hint: Use induction on $n$ and the fact that $\int_0^\infty e^{-x^2} dx = \dfrac{\sqrt{\pi}}{2}$ (Euler's formula).

19. Show that for $a > 0$, $\int_{-\infty}^\infty e^{-x^2} \cos(ax)\, dx = \sqrt{\pi} e^{-a^2/4}$.

Hint: Use the problem above.

20. Let $M$ be an arbitrary metric space and $f : \mathbb{R} \times M \to \mathbb{R}$ a function. Assume that $f(-, t)$ is a measurable function for each $t \in M$ and $f(x, -)$ is a continuous function for each $x \in \mathbb{R}$. Assume also that there exists an integrable function $g$ such that for each $t \in M$ we have $|f(x, t)| \leq g(x)$ for almost all $x \in \mathbb{R}$. Show that the function $F : M \to \mathbb{R}$ defined by

$$F(t) = \int_{\mathbb{R}} f(x, t) \ dx$$

is a continuous function.

## 2.8 Banach–Tarski Paradox

Consider the closed unit ball in three dimensions,

$$B = \{(x, y, z) \in \mathbb{R}^3 : \quad x^2 + y^2 + z^2 \leq 1\}.$$

Can $B$ be decomposed into a finite number of parts and reassembled into two balls identical to the original $B$? Our first response will be "No!" (Figure 2.28).

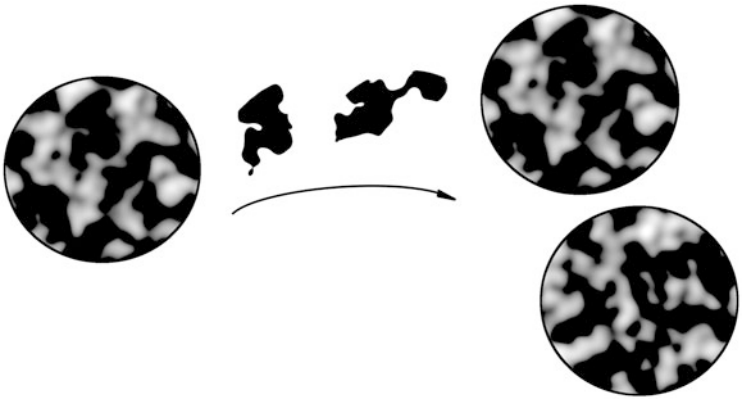


Figure 2.28: Banach–Tarski paradox

The intuition one has about volumes of everyday objects in the physical world makes us think this way. Doubling the ball by dividing it into parts and moving the parts around by rigid motions (i.e., rotations and translations) without any stretching, bending, or adding new points seems to be impossible. However, perhaps the intuition we rely on is not applicable to some subsets in the world of $\mathbb{R}^3$. In fact, in a paper published in 1924, Stefan Banach and Alfred Tarski gave a construction of such a paradoxical decomposition. The ordinary "volume" of a subset of $\mathbb{R}^3$ or $\mathbb{R}^n$ is its three-dimensional or $n$-dimensional Lebesgue measure, respectively. It is possible that not all subsets of $\mathbb{R}^n$ are measurable; in fact the pieces in the Banach–Tarski paradox are not measurable. Therefore, we can be assured that the rules concerning volume are actually not violated here. We will not prove the Banach–Tarski decomposition in this section. A proof of it and its consequences are given in a book by Wagon [53] in detail. In the following we answer the question of whether or not all subsets of $\mathbb{R}^n$ are measurable. For this we need to recall the axiom of choice and the definition of measure.

### Axiom of Choice

The Axiom of Choice appears in many proofs in mathematics. For a long time mathematicians discussed among themselves about the use of Axiom of Choice and seemingly contradictory results it produced. Now most seem to agree that

the advantages of the Axiom of Choice outweigh the disadvantages.
Suppose $I$ is an index set and $\{A_i\}_{i\in I}$is a collection of nonempty sets. We call $\psi$ a *choice function* if $\psi : I \to \bigcup_{i\in I} A_i$ is such that $\psi(i) \in A_i$.

**The Axiom of Choice:** For every collection of nonempty sets there exists a choice function.

Note that the Axiom of Choice (AC) is a nonconstructive assertion of existence. It postulates the existence of certain objects without giving any indication of how to find these objects. There are various logically equivalent statements such as Hausdorff Maximality Principle, Zorn's Lemma, and Well-Ordering Principle, see the end of Section (1.1) for more information.

## Construction of a Non-Measurable Set

Are all subsets of $\mathbb{R}^n$ measurable? In this section we answer this question for $n = 1$ by constructing a subset of $\mathbb{R}$ which is not measurable. The construction of a non-measurable set uses the Axiom of Choice. First we recall the concept of measure on $\mathbb{R}$.

**Definition 102.** We call a non-negative function $m$ a *measure* on the family of $E \subset \mathbb{R}$ if

a) $m(E) = b - a$ if $E = [a, b],\ a \le b$.

b) $m(E) = \sum_{n=1}^{\infty} m(E_n)$, whenever $E = \bigcup_{n=1}^{\infty} E_n$ and the sets $E_n$ are disjoint.

c) $m(E + h) = m(E)$ for every $h \in \mathbb{R}$.

The construction of a non-measurable set $\mathcal{M}$ rests on a simple equivalence relation among real numbers in $[0, 1]$. We say $x$ is equivalent to $y$, and write this as

$$x \sim y \quad \text{whenever} \quad x - y \quad \text{is rational.}$$

Note that this is an equivalence relation since the following properties hold:

- $x \sim x$ for every $x \in [0, 1]$.
- if $x \sim y$, then $y \sim x$.
- if $x \sim y$ and $y \sim z$, then $x \sim z$.

Two equivalent classes are either disjoint or coincide and furthermore $[0, 1]$ is the disjoint union of these equivalence classes $E_\alpha$. In other words,

$$[0, 1] = \bigcup_{\alpha} E_\alpha.$$

Now we construct a set $\mathcal{M}$ by choosing exactly one element $x_\alpha$ from each $E_\alpha$, and setting $\mathcal{M} = \{x_\alpha\}$. Note that the construction of the set $\mathcal{M}$ is possible because of the Axiom of Choice.

**Theorem 119** (G. Vitali: Non-measurable set)**.** *The set $\mathcal{M}$ is not measurable.*

*Proof.* Assume that $\mathcal{M}$ is measurable. Let $r_1, r_2, r_3, \ldots$ be an enumeration of all rationals in $[-1, 1]$, and consider the translates $\mathcal{M}_j = \mathcal{M} + r_j$ for $j = 1, 2, 3, \ldots$. First, observe that the sets $\mathcal{M}_j$ are disjoint, because of the fact that $\mathcal{M}$ contains only one representative of each equivalence class. Moreover each $\mathcal{M}_j$ by construction is contained in $[-1, 2]$. Finally, if $x \in [0, 1]$, then $x \sim x_\alpha$ for some $\alpha$, and therefore $x - x_\alpha = r_j$ for some $j$. Thus we have:

$$[0, 1] \subset \bigcup_{j=1}^{\infty} \mathcal{M}_j \subset [-1, 2].$$

The inclusions above yield:

$$m([0, 1]) = 1 \leq \sum_{j=1}^{\infty} m(\mathcal{M}_j) \leq 3 = m([-1, 2]).$$

Since $\mathcal{M}_j$ is a translate of $\mathcal{M}$, we must have

$$m(\mathcal{M}) = m(\mathcal{M}_j) \quad \text{for all} \quad j.$$

Consequently

$$1 \leq \sum_{j=1}^{\infty} m(\mathcal{M}) \leq 3.$$

This is the desired contradiction, since neither $m(\mathcal{M}) = 0$ nor $m(\mathcal{M}) > 0$ is possible. Here $\mathcal{M}$ is non-measurable. □

In the following we present two examples concerning Banach–Tarski type decomposition; both examples depend on the Axiom of Choice. Recall that two subsets of the plane are said to be congruent if one can be made to coincide precisely with the other using only translations and rotations in the plane.

In the following examples, the basic idea is divide an object $X$ into a finite number of disjoint parts and then "rearrange" them into a new object, say $Y$. In this case we say $X$ is equivalent by *finite decomposition* to $Y$. Note that rearranging a given set means that the set in its initial position is congruent to the set in its final position.

**Example 94.** We claim that $\mathbb{N}$ is equivalent by finite decomposition to the natural numbers with a one element "hole" in them. To this end, first create two subsets of $\mathbb{N}$:

- $A = \{1, 2, 3, 4, 6, 7, \ldots\}$ (non-multiples of 5).
- $B = \{5, 10, 15, \ldots\}$ (all multiples of 5).

Notice that

$$A \cap B = \emptyset \quad \text{and} \quad A \cup B = \mathbb{N}.$$

Next, we shift $B$ toward infinity by 5 units (we remove 5 from $\mathbb{N}$), and thus produce the new set

$$B^* = \{10, 15, 20, \dots\}.$$

We still have:

$$A \cap B^* = \emptyset \quad \text{and} \quad A \cup B^* = \mathbb{N} \setminus \{5\}.$$

Moreover, $B$ and $B^*$ are congruent. Thus $\mathbb{N}$ and $\mathbb{N} \setminus \{5\}$ are equivalent by finite decomposition.

The Banach–Tarski construction is similar to an older construction due to G. Vitali, which concerns area rather than the volume.

**Example 95** (Doubling Area)**.** Let $L_\theta$ denote the line segment in $\mathbb{R}^2$ represented as

$$L_\theta = \{(r, \theta) : \ 0 < r \leq 1\}.$$

Observe that union of all such segments is the punctured unit disc $B'$, i.e., $B'$ is the unit disc in $\mathbb{R}^2$ with the origin removed. Next we define an equivalence relation:
$L_\theta$ and $L_\phi$ belong to the same equivalence class if

$$\theta - \phi \qquad \text{is a rational multiple of } \pi.$$

Consider a set $S$ that is the union of a set of $L_\theta$ containing exactly one element from each equivalence class (notice how we used the Axiom of Choice). Let

$$S_n = \{L_{\theta + 2\pi r_n} : \ L_\theta \in S\}.$$

Note that each $S_n$ is obtained from $S$ by a rotation about the origin through an angle $2\pi r_n$, where $r_n \in [0, 1)$ is a sequence $r_1, r_2, \dots$ of rationals. Note that the $S_n$'s are disjoint and

$$B' = \bigcup_n S_n.$$

Now we split $B'$ into two sets as follows:

$$C = \bigcup_n S_{2n} \quad \text{and} \quad D = \bigcup_n S_{2n+1}.$$

Observe that each $S_{2n}$ can be rotated to $S_n$, and the union of $S_n$ give $B'$. Similarly, each $S_{2n+1}$ can be rotated to $S_n$ and the union of $S_n$ gives $B'$ again. Thus $B'$ can be split into a countable set of disjoint pieces which can be rotated and translated to form disjoint sets whose union is two copies of $B'$.

Although the Banach–Tarski Decomposition clearly relies on the Axiom of Choice, in 1991, Pawlikowski [39] showed that the Banach–Tarski Theorem is implied by the famous the Hahn–Banach Theorem of functional analysis. Thus a weakened version of the Axiom of Choice will do. If you are asking the question what makes the Axiom of Choice true or false, notice that when we accept the Axiom of Choice, we declare our intention to treat certain mathematical objects as they exist regardless of if we can find examples of these objects. Of course, one can always refer to the inscription written on the tombstone of David Hilbert: "Wir Müssen Wissen, Wir Werden Wissen" (We Must Know, We Will Know).

# Chapter 3

# Complex Analysis

> Neither the true nor the false roots are always real; sometimes they are imaginary; that is, while we can always imagine as many roots for each equation as I have assigned, yet there is not always a definite quantity corresponding to each root we have imagined.
>
> *Descartes* 1637

## 3.1 The Complex Plane

### Complex Numbers

Our idea of what a number is has evolved considerably throughout history. The ancients thought of numbers as extensions in space; we have looked at cardinalities, ratios, and invented negative numbers which are useful for describing debts, backward motion, etc. One more modern concept of what numbers are involves functions known as *polynomials*.

A *polynomial* is a sum of multiples of powers of a variable $x$, that is

$$p(x) = a_0 + a_1 x + \cdots + a_n x^n.$$

A number $\alpha$ is a *root* of a polynomial $p(x)$ if $p(\alpha) = 0$, meaning if

$$a_0 + a_1 \alpha + \cdots + a_n \alpha^n = 0.$$

Every real number $\alpha$ is a root of a polynomial, namely $p(x) = x - \alpha$. Thus, one idea for what it means to be a number is being "a root of a polynomial." However, not every polynomial has a real root. For example, there is no real number $x$ such that $x^2 + 1 = 0$. It turns out that we can fix this by inventing a single new number $i$ with the property that $i^2 = -1$.

A *complex number* is a sum $z = a + ib$ of a real number $a$ and $i$ times a real number $b$; $a$ is called the *real part* of $z$ denoted by $\operatorname{Re} z$, while $b$ is called the

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

A. G. Aksoy, *Fundamentals of Real and Complex Analysis*, Springer Undergraduate Mathematics Series, https://doi.org/10.1007/978-3-031-54831-4_3

*imaginary part* of $z$ and denoted by $\operatorname{Im} z$. The term "imaginary" is unfortunate since it misleadingly suggests that these numbers are somehow fake or unreal.

Where real numbers can be understood as lengths of segments, complex numbers can be identified with points of a plane (Figure 3.1):

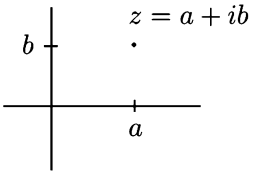



Figure 3.1: Cartesian representation

- We can identify a complex number either by using Cartesian coordinates (Figure 3.2) $z = a + ib$ or by using *polar coordinates*

$$z = re^{i\theta} = r\cos\theta + ir\sin\theta$$

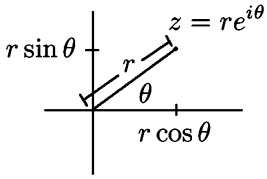



Figure 3.2: Polar representation

We write $e^{i\theta}$ as shorthand for $\cos\theta + i\sin\theta$ and $z = re^{i\theta}$ is called the polar form of a complex number. Here $r = |z| = \sqrt{x^2+y^2}$ and for $z = 0$, we can choose $\theta$ arbitrarily; for $z \neq 0$, the angle $\theta$ is not unique. Because of the periodic nature of $\sin\theta$ and $\cos\theta$, $\theta$ is only determined up to an integer multiple of $2\pi$. We call any value of $\theta$ with $z = re^{i\theta}$ *an argument* of $z$. In Section 3.4 we will examine the series expansion of the exponential function $e^z$ that properly justifies the notation $e^{i\theta} = \cos\theta + i\sin\theta$. Note that if we extend the series definition of the exponential function to complex numbers, we have for any real $\theta$,

$$\begin{aligned} e^{i\theta} &= 1 + i\theta + \frac{1}{2!}(i\theta)^2 + \frac{1}{3!}(i\theta)^3 + \frac{1}{4!}(i\theta)^4 + \frac{1}{5!}(i\theta)^5 + \cdots \\ &= \left(1 - \frac{1}{2!}\theta^2 + \frac{1}{4!}\theta^4 - \cdots\right) + i\left(\theta - \frac{1}{3!}\theta^3 + \frac{1}{5!}\theta^5 - \cdots\right) \\ &= \cos\theta + i\sin\theta. \end{aligned}$$

- The following are some special complex numbers (Figure 3.3). For $\theta \in \mathbb{R}$,

$$e^{i0} = 1 \quad \text{and} \quad e^{i\theta} = 1 \iff \theta = 2k\pi, \ \ (k \in \mathbb{Z}),$$

$$e^{i\pi} = -1 \quad \text{and} \quad e^{i\theta} = -1 \Longleftrightarrow \theta = (2k+1)\pi, \ \ (k \in \mathbb{Z}),$$
$$e^{i\pi/2} = i \quad \text{and} \quad e^{i\theta} = i \Longleftrightarrow \theta = \frac{1}{2}(2k+1)\pi, \ \ (k \in \mathbb{Z}).$$

Note that *Euler's formula*

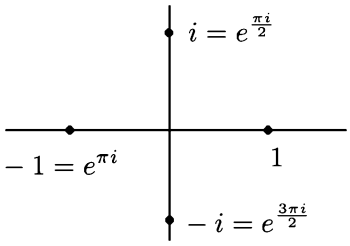



Figure 3.3: Some complex numbers

$$e^{\pi i} + 1 = 0$$

connects the five most important constants of mathematics, symbolizing the four major branches of classical mathematics:

a) Arithmetic represented by 0 and 1

b) Algebra by $i$

c) Geometry by $\pi$

d) Analysis by $e$

This formula also contains three of the most important operations: addition, multiplication, and exponentiation.

## Geometry in the Complex Plane

Many subsets of $\mathbb{C}$ that appear repeatedly later on are described below.

- For $z, w \in \mathbb{C}$, the *line segment* with end points $z$ and $w$ is given by

$$[z, \ w] = \{(1-t)z + tw : \quad 0 \leq t \leq 1\}.$$

- The *real axis* can be given by any of the following equations:

  a) $\operatorname{Im} z = 0$.

  b) $z = \overline{z}$.

  c) $|z - i| = |z + i|$.

  Similarly, the imaginary axis can be described as $\operatorname{Re} z = 0$ or $|z-1| = |z+1|$ (Figure 3.4).

- The *circle* with center $a \in \mathbb{C}$ and radius $r > 0$ has the equation $|z-a| = r$.
- The *open disc* with center $a \in \mathbb{C}$ and radius $r > 0$ is defined to be

$$D(a;r) := \{z \in \mathbb{C} : \quad |z-a| < r\},$$

and the *closed disc* with center $a$ and radius $r > 0$ is

$$\overline{D(a;r)} := \{z \in \mathbb{C} : \quad |z-a| \leq r\}.$$

- A set of the form $\{z \in \mathbb{C} : \quad s < |z-a| < r\} \quad \text{for} \quad (0 \leq s < r)$ is called an open *annulus*:

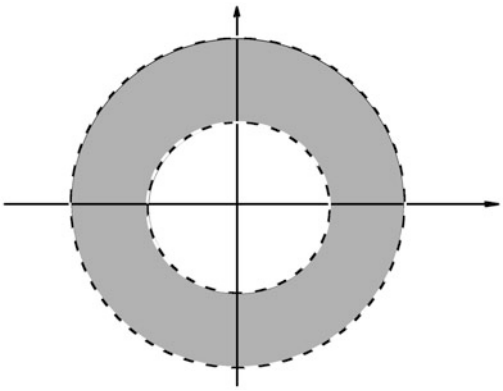


Figure 3.4: Open annulus

- The set

$$\prod\nolimits^{+} := \{z \in \mathbb{C} : \operatorname{Im} z > 0\}$$

is called the open *upper half-plane* (Figure 3.5):

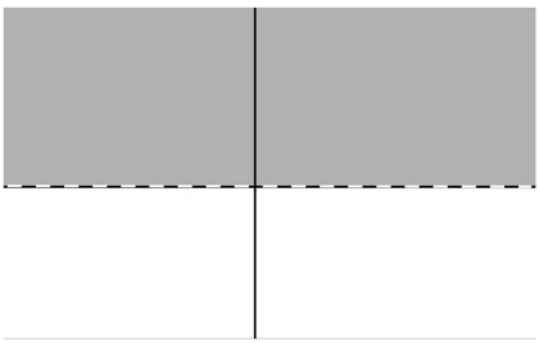


Figure 3.5: Upper half-plane

Similarly,

$$\overline{\prod} := \{z \in \mathbb{C} : \operatorname{Im} z < 0\}$$

will be the open *lower half-plane*. To obtain the closed upper half-plane simply change $>$ to $\geq$. Other half-planes are defined analogously.

- The set (Figure 3.6)

$$S_{\alpha,\beta} := \{z \in \mathbb{C} : \quad 0 \neq z = |z|e^{i\theta} \text{ with } \alpha < \theta < \beta\}$$

defines a *sector*.

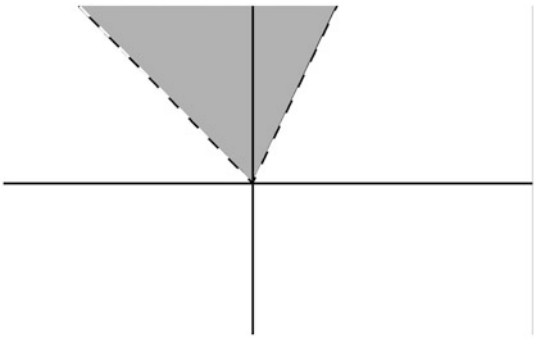


Figure 3.6: A sector

- Every complex number $z = a + ib$ has a *complex conjugate* $\overline{z} = a - ib$ with the property that $z\overline{z} \in \mathbb{R}$; the map $f(z) = \overline{z}$ is reflection across the real axis (Figure 3.7).

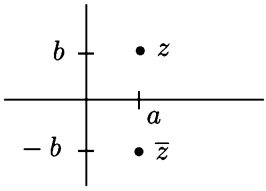



Figure 3.7: Complex conjugate

- Adding a complex number $w$, i.e., $f(z) = z + w$ translates the plane by moving 0 to $w$ (Figure 3.8):

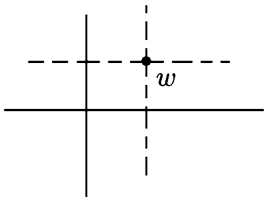



Figure 3.8: Moving 0 to $w$

- Multiplication by a positive real number $r$ scales the plane by a factor of $r$ (Figure 3.9),

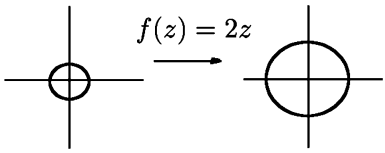



Figure 3.9: Multiplication by 2

while multiplication by a negative real number $r$ reflects through the origin and scales by $|r|$ (Figure 3.10).

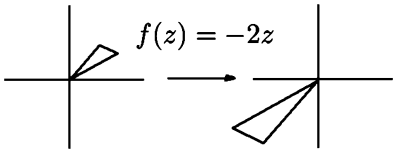



Figure 3.10: Multiplication by $-2$

- Multiplication by a number of the form $e^{i\theta}$ rotates the plane by an angle of $\theta$ about the origin (Figure 3.11).

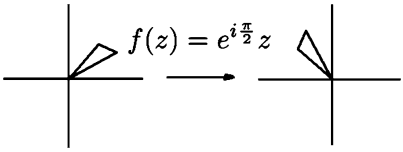



Figure 3.11: Multiplication by $e^{i\theta}$

In order to understand the above property we examine the complex exponential function in detail later in Section 3.4.

## Arithmetic of Complex Numbers

The real part of a complex number $z = a + ib$ is $a$ and is denoted by $\operatorname{Re} z$, and the number $b$ is called the imaginary part of $z$ and is denoted by $\operatorname{Im} z$. For example, $\operatorname{Re}(3 + 2i) = 3$ and $\operatorname{Im}(1 - 5i) = -5$. Two complex numbers are considered equal if and only if their real and imaginary parts are equal, that is,

$$a + bi = c + di \quad \text{if and only if} \quad a = c \quad \text{and} \quad b = d.$$

Complex numbers are added, subtracted, and multiplied in accordance with the standard rules of algebra but with $i^2 = -1$:

$$\begin{aligned}(a + bi) + (c + di) &= (a + c) + (b + d)i,\\ (a + bi) - (c + di) &= (a - c) + (b - d)i,\\ (a + bi)(c + di) &= (ac - bd) + (ad + bc)i.\end{aligned}$$

Also note that if $b = 0$, then the multiplication formula simplifies to

$$a(c + di) = ac + adi.$$

The product of a complex number $z = a + bi$ and its *conjugate* $\overline{z} = a - ib$ is a non-negative real number:

$$z\overline{z} = (a + bi)(a - bi) = a^2 - abi + bai - b^2 i^2 = a^2 + b^2.$$

Note that complex conjugate of $z$ in the polar form $z = re^{i\theta}$ is $\overline{z} = re^{-i\theta}$. You will recognize that

$$|z| = \sqrt{z\overline{z}} = \sqrt{a^2 + b^2}$$

is the length of the vector corresponding to $z$; we call this length the *modulus* (or absolute value) of $z$ and denote it by $|z|$. For example the modulus of $z = -4-5i$ is $|z| = \sqrt{(-4)^2 + (-5)^2} = \sqrt{41}$.
If $z \neq 0$, then the *reciprocal* (or multiplicative inverse) of $z$ is denoted by $1/z$ (or $z^{-1}$) and is defined by the property

$$\left(\frac{1}{z}\right) z = 1.$$

This equation has a unique solution for $1/z$, which can be obtained by multiplying both sides by $\overline{z}$ and using the fact $z\overline{z} = |z|^2$. This yields the expression

$$\frac{1}{z} = \frac{\overline{z}}{|z|^2}.$$

If $z_2 \neq 0$, then the *division* $z_1/z_2$ is defined as the product of $z_1$ and $1/z_2$, and thus

$$\frac{z_1}{z_2} = \frac{\overline{z_2}}{|z_2|^2} z_1 = \frac{z_1 \overline{z_2}}{|z_2|^2}.$$

Note that the expression on the right-hand side of the above equation can be viewed as both the numerator and denominator of $z_1/z_2$ multiplied by $\overline{z_2}$.

**Example 96.** Let $z_1 = 3 + 4i$ and $z_2 = 1 - 2i$. To find $z_1/z_2$ we will multiply the numerator and denominator of $z_1/z_2$ by $\overline{z_2}$.

$$\begin{aligned} \frac{z_1}{z_2} = \frac{z_1 \overline{z_2}}{z_2 \overline{z_2}} &= \frac{3+4i}{1-2i} \frac{1+2i}{1+2i} \\ &= \frac{3 + 6i + 4i + 8i^2}{1 - 4i^2} \\ &= \frac{-5 + 10i}{5} \\ &= -1 + 2i. \end{aligned}$$

The following identities hold for all $z, w \in \mathbb{C}$:

a) $\overline{\overline{z}} = z$.

b) $\operatorname{Re} z = \dfrac{z + \overline{z}}{2}$ and $\operatorname{Im} z = \dfrac{z - \overline{z}}{2i}$.

c) $|\overline{z}| = |z|$.

d) $|z|^2 = z \cdot \overline{z}$.

e) $\overline{zw} = \overline{z} \cdot \overline{w}$.

f) $\overline{z + w} = \overline{z} + \overline{w}$.

**Remark.** The set of complex numbers, denoted $\mathbb{C}$, form a field; in fact, the complex numbers are an *algebraically complete* field, meaning that every polynomial with complex coefficients has a complex root. In fact, we have the following important theorem.

**Theorem 120.** *(Fundamental Theorem of Algebra) Every polynomial of degree $n$ with complex coefficients has $n$ complex roots, counted with multiplicity.*

For the proof we refer the reader to Theorem 137 in Section 3.8 below.
The Fundamental Theorem of Algebra is an existence theorem and does not tell you a method for finding zeros. For polynomials of degree 2, quadratic formula gives the zeros explicitly. Similar but more complicated formulas exist for polynomials of degree 3 and 4. No such formulas exist for polynomials of degree 5 and higher. The Fundamental Theorem of Algebra fails in $\mathbb{R}$. This difference accounts for the differences between operators defined real and complex vector spaces.

## Products and Powers; de Moivre's Theorem

There are certain polynomials which occur repeatedly in complex analysis. It is essential to became familiar with these polynomials and the location of their roots. First notice that the polar form of a nonzero complex number of the *exponential form*

$$z = re^{i\theta} = r\cos\theta + ir\sin\theta$$

enables us to multiply two complex numbers, say $z = r(\cos\theta + i\sin\theta)$ and $w = \rho(\cos\psi + i\sin\psi)$. We have

$$z\,w = r\,\rho[(\cos\theta + i\sin\theta)\,(\cos\psi + i\sin\psi)] = r\rho\cos(\theta+\psi) + ir\rho\sin(\theta+\psi)$$

by the standard trigonometric formulas, thus

$$z\,w = r\,\rho\,e^{(\theta+\psi)}.$$

This implies in particular that $|z\,w| = |z|\,|w|$. Let $z = re^{i\theta}$ and $n$ be a natural number, then from the above and a routine proof by induction, we obtain

$$z^n = r^n\,e^{in\theta}.$$

In the special case $r = 1$ we obtain *de Moivre's theorem*:

$$(\cos\theta + i\sin\theta)^n = \cos n\theta + i\sin n\theta, \quad (\theta \in \mathbb{R})$$

which is a compact expression for a family of trigonometric identities. For example, in view of the binomial theorem one can show:

$$\cos 5\theta = \cos^5\theta - 10\cos^3\theta\sin^2\theta + 5\cos\theta\sin^4\theta.$$

Also for $z = re^{i\theta} \neq 0$, we have $\dfrac{1}{z} = r^{-1}e^{-i\theta}$.

**Example 97.** If $z = 1+i$, then $|z| = \sqrt{2}$ and $z$ lies in the first quadrant on the line $y = x$. Hence $z = \sqrt{2}\,e^{\,i\frac{\pi}{4}}$ and $\dfrac{1}{z} = \dfrac{1}{\sqrt{2}}e^{\,-i\frac{\pi}{4}}$.

The fundamental theorem of algebra above guarantees that a polynomial of degree $n$ with coefficients in $\mathbb{C}$ has $n$ roots (not necessarily distinct, of course). Let us consider the equation $z^n = 1$; using polar form of $z = re^{i\theta}$, we must have:

$$z^n = 1 \Longleftrightarrow r^n e^{ni\theta} = 1 \Longleftrightarrow r = 1, \quad \cos n\theta + i\sin n\theta = 1.$$

Thus the distinct roots of the equation $z^n = 1$ are given by

$$e^{\frac{2k\pi i}{n}} \quad \text{for} \quad k = 0, 1, \ldots, n-1.$$

These numbers are known as the $n$th *roots of unity*. Observe that the roots of $z^n = 1$ lie at the vertices of a regular $n$-gon centered at 0 with one vertex at 1 (Figure 3.12).

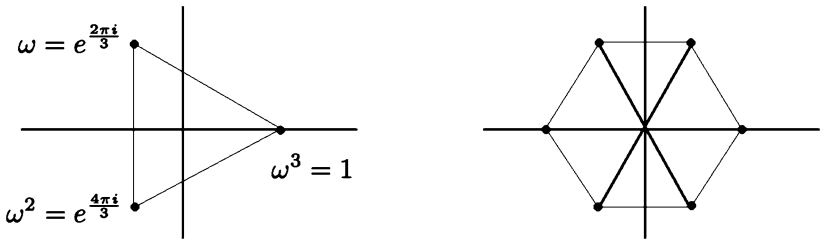



Figure 3.12: Left: cube roots of unity, Right: sixth roots of unity

The first picture on the left is the cube roots of unity and

$$z^3 = 1 \Longleftrightarrow z = \omega^3 = 1, \quad \text{or} \quad z = \omega = e^{\frac{2\pi i}{3}} \quad \text{or} \quad z = \omega^2 = e^{\frac{4\pi i}{3}}.$$

The picture on the right represents the sixth roots of unity.
Two special cases are worth mentioning. We have:

$$z^4 = 1 \Longleftrightarrow (z^2 - 1)(z^2 + 1) = 0 \Longleftrightarrow z = \pm 1 \quad \text{or} \quad z = \pm i$$

and the geometric identity

$$(1 - z)(1 + z + z^2 + \cdots + z^k) = 1 - z^{k+1}$$

is valid for all $z \in \mathbb{C}$ and for all natural numbers $k$. For example, if we set $k = 2$, we obtain

$$(1 - z)(1 + z + z^2) = 1 - z^3,$$

which gives the cube roots of unity which we calculated before. Therefore, the roots of the equation $1 + z + z^2 = 0$ are exactly the non-real cube roots of unity, that is $\omega$ and $\omega^2$.

## Inequalities

The following are some important inequalities that link modulus and addition.

**Proposition 40.** *For all $z$ and $w$ in $\mathbb{C}$,*

*a)* $|Re\, z| \leq |z|$ *and* $|Im\, z| \leq |z|$.

*b)* $|z + w| \leq |z| + |w|$ *(the triangle inequality).*

*c)* $\big||z| - |w|\big| \leq |z + w|$.

*Proof.* Part a) is clear since $|z|^2 = (\mathrm{Re}\, z)^2 + (\mathrm{Im}\, z)^2$ and $|z| \geq 0$. We leave it to the reader to prove part b). For c) all we need is that the following two inequalities hold:

$$|w| - |z| \leq |z + w| \quad \text{and} \quad |z| - |w| \leq |z + w|.$$

Using the triangle inequality we have

$$|w| = |z + w - z| \leq |z + w| + |-z|$$

and

$$|z| = |z + w - w| \leq |z + w| + |-w|,$$

so c) follows. □

**Remark.** All of the inequalities given above are between real numbers ($\operatorname{Re} z$, $|z|$ etc.). Note that there is no meaning assigned to an inequality $z \leq w$. In fact it is not possible to define an ordering on $\mathbb{C}$. Suppose such an ordering exists, then

$$i \geq 0 \quad \text{implies} \quad (i)(i) \geq 0, \quad -1 \geq 0$$

or

$$i \leq 0 \quad \text{implies} \quad (-i)(-i) \geq 0, \quad -1 \geq 0,$$

contradicting the order axioms.

## The Riemann Sphere

The extended complex plane, which consists of $\mathbb{C}$ and the point at $\infty$, has a convenient interpretation which we now explain. Consider the Euclidean space $\mathbb{R}^3$ with coordinates $(x, y, z)$ where the $xy$-plane is identified with $\mathbb{C}$. We denote by $\mathbb{S}$ the sphere centered at $(0, 0, 1/2)$ of radius $1/2$. In other words,

$$\mathbb{S} = \{(x, y, u) : \quad x^2 + y^2 + (u - 1/2)^2 = 1/4\}.$$

This sphere is the *Riemann sphere* which touches the plane $\mathbb{C}$ at the point $(0, 0, 0)$. Let $N = (0, 0, 1)$ be the north pole of the sphere. Given any point $z' = (x, y, u)$ on $\mathbb{S}$ different from the north pole, the line joining $N$ to $z'$ intersects the $xy$-plane at a single point denoted by $z = x + iy$. We call $z$ the *stereographic projection* of $z'$ (Figure 3.13).

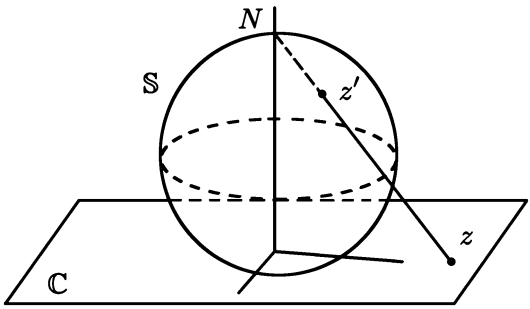



Figure 3.13: Stereographic projection

Conversely, given any point $z \in \mathbb{C}$, the line joining $N$ to $z$ intersects the sphere at $N$ and another point which we call $z'$. Thus, there is a bijective correspondence between points of $\mathbb{S} \setminus N$ and the complex plane $\mathbb{C}$. The problem of the north pole being left out can be handled by adding an extra point $\infty$ to $\mathbb{C}$. We define the *extended complex plane*

$$\tilde{\mathbb{C}} := \mathbb{C} \cup \infty.$$

We then have the natural correspondence between $\tilde{\mathbb{C}}$ and $\mathbb{S}$ by

$$z = x + iy \longleftrightarrow z' = \left( \frac{x}{1+r^2}, \frac{y}{1+r^2}, \frac{r^2}{1+r^2} \right)$$

where $r^2 = x^2 + y^2$ and under this bijection $\infty \longleftrightarrow (0,0,1) = N$.

## Exercises

1. Prove that $\dfrac{1}{i} = -i$ and that $\dfrac{1}{i+1} = \dfrac{1-i}{2}$.

2. Let $z_1 = 2 + i$, $z_2 = 3 - 2i$ and $z_3 = -\dfrac{1}{2} + \dfrac{\sqrt{3}}{2}\, i$. Evaluate:

   a) $|3z_1 - 4z_2|$, b) ${z_1}^3 - 3{z_1}^2 + 4z_1 - 8$, c) $\dfrac{z_1}{z_2}$, d) $(\overline{z_3})^4$.

3. Find real numbers $x$ and $y$ such that

$$3x + 2iy - ix + 5y = 7 + 5i.$$

4. Find the real and imaginary parts of the following where $z = x + iy$:

   a) $\dfrac{1}{z^2}$ b) $\dfrac{1}{3z+2}$.

5. If $a$ is real and $z$ is complex, prove that $\mathrm{Re}(az) = a\,\mathrm{Re}(z)$ and that $\mathrm{Im}(az) = a\,\mathrm{Im}(z)$. Generally, show that $\mathrm{Re} : \mathbb{C} \longrightarrow \mathbb{R}$ is a linear map, that is,

$$\mathrm{Re}(az + bw) = a\,\mathrm{Re}\,z + b\,\mathrm{Re}\,w$$

for $a, b$ real, $z, w$ complex.

6. Prove that, for any integer $k$,

$$i^{4k} = 1, \qquad i^{4k+1} = i, \qquad i^{4k+2} = -1, \qquad i^{4k+3} = -i.$$

Show how this result gives a formula for $i^n$ for all $n$ by writing $n = 4k + j$, $0 \le j \le 3$.

7. Sketch the sets of points that satisfy the following equations:

$$\text{a) } |z| = 2, \qquad \text{b) } |z - i| = |z + i|, \qquad \text{c) } |2z - 4i| < 1,$$

$$\text{d) } 2 > \operatorname{Re} z > -3, \qquad \text{e) } (\operatorname{Im} z)^2 \leq \operatorname{Re} z.$$

8. Show that the field of complex numbers $\mathbb{C}$ is not an ordered field.

9. Let $p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$ be a polynomial with real coefficients $a_0, a_1, a_2, \ldots, a_n$. Prove that if $z$ is a solution of the equation $p(x) = 0$, then so is $\overline{z}$.

10. Solve $z^8 = 1$ for $z$.

11. Find each of the indicated roots and locate them graphically:

$$\text{a) } (-1 + i)^{\frac{1}{3}}, \qquad \text{b) } (-2\sqrt{3} - 2i)^{\frac{1}{4}}.$$

12. Express $\cos 3\theta$ in terms of $\cos\theta$ and $\sin\theta$ using de Moivre's formula.

13. For $z \neq 1$ verify the identity

$$1 + z + z^2 + \cdots + z^n = \frac{1 - z^{n+1}}{1 - z}.$$

14. Use the problem 13 above to establish that

$$1 + \cos\theta + \cos 2\theta + \cdots + \cos n\theta = \frac{1}{2} + \frac{\sin(n + \frac{1}{2})\theta}{2\sin\frac{1}{2}\theta}$$

for $0 < \theta < 2\pi$. This result is known as *Lagrange's identity* and is useful in the theory of Fourier series.

## 3.2 Holomorphic Functions

Complex analysis is the study of holomorphic (sometimes called analytic) functions. Holomorphic means almost the same thing as differentiable, but there is an important distinction between the two concepts caused by the role played by open sets. In studying differentiable functions of a real variable, we consider functions defined on intervals. For a complex-valued function, we need to select the domains of our definition as an open set.

**Definition 103.** Let $f : G \subset \mathbb{C} \to \mathbb{C}$ be a function from an open set $G$ of complex numbers to the complex numbers. We say $f$ is *differentiable* at $z \in G$ if

$$\lim_{h \to 0} \frac{f(z + h) - f(z)}{h}$$

exists (independently of the manner in which $h$ approaches 0) for each $z \in G$. This limit is denoted by $f'(z)$ and is called the complex derivative.

**Definition 104.** A complex-valued function $f$ which is differentiable at every point of an open set $G$ is said to be *holomorphic* in $G$ and we use $H(G)$ to denote the set of all functions holomorphic in $G$.

A complex-valued function $f$ is said to be *holomorphic at a point* $a \in \mathbb{C}$ if there exists $r > 0$ such that $f$ is defined and is holomorphic in $D(a; r)$, where $D(a; r)$ is the open disc defined as $D(a; r) := \{z \in \mathbb{C} : |z - a| < r\}$. Note that being holomorphic at a point $a$ is stronger than simply being differentiable at $a$.

Note that $G$ is open, thus, given $z \in G$ there exists $r > 0$ such that $D(z, r) \subset G$. That is, $z + h \in G$ whenever $|h| < r$. This ensures that in the definition of $f'(z)$, the point $z + h$ is free to approach $z$ from any direction as $h \to 0$. For $f$ to be differentiable at $z$, it is necessary that the quotient $\dfrac{f(z+h) - f(z)}{h}$ tends to a limiting value independent of the way $h \to 0$. In particular, $f$ cannot be differentiable at $z$ if the quotient $\dfrac{f(z+h) - f(z)}{h}$ has different limiting values when $h \to 0$ from different directions. For example, $f(z) = \overline{z}$ is not holomorphic since,

$$\lim_{h\to 0} \frac{\overline{(z+h)} - \overline{z}}{h} = \lim_{h\to 0} \frac{\overline{h}}{h} = \begin{cases} 1 & \text{if } h \text{ is real} \\ -1 & \text{if } h \text{ is purely imaginary.} \end{cases}$$

The idea of restricting $h$ to real and imaginary values can be used to obtain a necessary but not sufficient condition for differentiability.

## Cauchy–Riemann Equations

In this section we will see that a differentiable function gives rise to a set of wonderful partial differential equations called the Cauchy–Riemann equations.

**Theorem 121.** *Let $f$ be a complex-valued function defined on an open set $G$ and let $f$ be differentiable at $z = x + iy \in G$. Let $f(z) = u(x, y) + iv(x, y)$. Then $u$ and $v$ have first-order partial derivatives (denoted by $u_x$, $u_y$, $v_x$, $v_y$) and these partial derivatives satisfy the Cauchy–Riemann equations:*

$$u_x = v_y \ , \ v_y = -v_x.$$

*Proof.* Since $f$ is differentiable at $z \in G$, $f'(z) = \lim\limits_{h\to 0} \dfrac{f(z+h) - f(z)}{h}$ exists. We calculate $f'(z)$ in two different ways, restricting $h$ to be real and purely imaginary.
Case 1: Suppose $h$ is real. Then writing

$$\begin{aligned} f(z+h) &= u(z+h) + iv(z+h) \\ &= u(x+h, y) + iv(x+h, y) \end{aligned}$$

in the definition of $f'(z)$, we have

$$\begin{aligned} f'(z) &= \lim_{h\to 0} \frac{f(z+h)-f(z)}{h} \\ &= \lim_{h\to 0} \left[ \frac{u(x+h,y)-u(x,y)}{h} + i\frac{v(x+h,y)-v(x,y)}{h} \right] \\ &= u_x + iv_x. \end{aligned}$$

Case 2: Suppose $h$ is purely imaginary (i.e., $h = ik$, $k \in \mathbb{R}$). Then,

$$f(z+h) = f(z+ik) = u(x,y+k) + iv(x,y+k)$$

and

$$\begin{aligned} f'(z) &= \lim_{\substack{h\to 0 \\ h=ik}} \left[ \frac{u(x,y+k)-u(x,y)}{ik} + i\frac{v(x,y+k)-v(x,y)}{ik} \right] \\ &= \frac{1}{i}u_y + v_y. \end{aligned}$$

Equating the two expressions for $f'(z)$

$$u_x + iv_x = -iu_y + v_y$$

and equating real and imaginary parts we obtain

$$\begin{aligned} u_x &= v_y \\ u_y &= -v_x. \end{aligned}$$

□

The converse of Theorem 121 is false as seen in the following example:

**Example 98.** Let $f(z) = f(x+iy) = 0$ if both $x$ and $y$ are zero and $f(z) = 1$ if neither $x$ nor $y$ is zero. Then $u_x = u_y = v_x = v_y = 0$. Thus, the Cauchy–Riemann equations hold. However, $f(z)$ is not differentiable at 0, because

$$\lim_{h\to 0} \frac{f(0+h)-f(0)}{h}$$

fails to exist if $h \to 0$ along the ray $\arg z = \pi/4$.

Note that the contrapositive of Theorem 121 is useful for proving non-differentiability, as shown in the following example.

**Example 99.** To show the complex function $f(z) = x+4iy$ is not differentiable at any point $z$, we identify the real and imaginary parts of $f(z)$:

$$\begin{aligned} u(x,y) = x, &\quad v(x,y) = 4y, \\ u_x = 1, &\quad v_x = 0, \\ u_y = 0, &\quad v_y = 4. \end{aligned}$$

In view of $u_x = 1 \neq v_y = 4$ the equations given in Theorem 121 cannot be simultaneously satisfied at any point $z$. Therefore $f$ is nowhere differentiable.

It turns out that the Cauchy–Riemann equations plus the extra assumption of continuity of the partial derivatives imply the converse of Theorem 121.

**Theorem 122.** *Let $z = x + iy \in G$, where $G \in \mathbb{C}$ is an open set, and let $f(z) = u(x, y) + iv(x, y)$. Assume $u(x, y)$ and $v(x, y)$ have continuous first-order partial derivatives in $G$ and they satisfy the Cauchy–Riemann equations at $z$. Then $f'(z)$ exists.*

The proof is technical, so we only state the result and refer the reader to [41], p. 59.

**Remark 68.** Complex multiplication can be interpreted as a linear map from $\mathbb{R}^2$ to $\mathbb{R}^2$ and thus as a $2 \times 2$ matrix as follows. Fix $w = a + ib$ a complex number, then for any other complex number $z = x + iy$, we have by complex multiplication:

$$wz = (a + ib)(x + iy) = (ax - by) + i(ay + bx).$$

Now, representing $z = \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2$ as a vector, we get

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax - by \\ bx + ay \end{bmatrix}.$$

Clearly, the matrix $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ represents multiplication by a complex number $a + bi$ if and only if $a_{11} = a_{22} = a$ and $a_{21} = -a_{12} = -b$.

Now, associate to each complex function $f(z) = u(x, y) + iv(x, y)$ the mapping $f(x, y) = (u(x, y), v(x, y))$ from $\mathbb{R}^2$ to $\mathbb{R}^2$. We know if $f$ is differentiable, the partials of $u$ and $v$ exist, and the Jacobian matrix of $f$ at some point $z_0 = (x_0, y_0)$ denoted by $J(x_0, y_0)$ is given as

$$J(x_0, y_0) = \begin{bmatrix} u_x(x_0, y_0) & u_y(x_0, y_0) \\ v_x(x_0, y_0) & v_y(x_0, y_0) \end{bmatrix}.$$

If the Cauchy–Riemann equations are satisfied, in others words when $u_x = v_y$ and $v_x = -u_y$, we see that the Jacobian represents multiplication by a complex number. It is also not difficult to see the determinant of the Jacobian:

$$\begin{aligned} \det J(x_0, y_0) &= u_x v_y - v_x u_y \\ &= u_x^2 + v_y^2 \\ &= |f'(z_0)|^2. \end{aligned}$$

## Elementary Properties of Holomorphic Functions

Let $G$ be a fixed open set, then

1) If $f, g \in H(G)$ and $\alpha \in \mathbb{C}$ then

$$\alpha f,\ f + g, \text{ and } fg \text{ (all defined pointwise)}$$

are holomorphic in $G$. Moreover,

$$\begin{aligned}(\alpha f)'(z) &= \alpha f'(z)\\ (f+g)'(z) &= f'(z)+g'(z)\\ (fg)'(z) &= f'(z)g(z)+g'(z)f(z).\end{aligned}$$

The usual rules hold.
2) If $f$ is holomorphic in $G$ and $g$ is holomorphic in $U$ where $U$ is an open set containing $f(G)$, then the composite function $g \circ f$ is defined as

$$\begin{aligned}(g\circ f)(z) &= g(f(z))\ \forall z \in G \text{ and}\\ (g\circ f)'(z) &= g'(f(z))f'(z).\end{aligned}$$

3) If $f$ is holomorphic in $G$ when $f(z) \neq 0$ for all $z \in G$, then $1/f(z)$ is holomorphic in $G$ and

$$\left(\frac{1}{f}\right)'(z) = \frac{-f'(z)}{(f(z))^2}.$$

The elementary properties listed above produce some preliminary holomorphic functions. In fact $f(z) = z$ and $f(z) = c$ (constant function) are certainly differentiable everywhere, thus any polynomial

$$p(z) = \sum_{n=0}^{N} c_n z^n,$$

where $c_n \in \mathbb{C}$, is holomorphic in $\mathbb{C}$. A rational function $\dfrac{p(z)}{q(z)}$ where $p(z)$ and $q(z)$ are polynomials and $q(z) \neq 0$ is holomorphic. For example $\dfrac{z}{1+z^2}$ is holomorphic at every point except $i$ and $-i$.

4) Holomorphy implies continuity. This follows from the fact that if $f'(z)$ exists then $\lim_{z\to z_0} f(z) = f(z_0)$ because

$$\lim_{z\to z_0}\left[\frac{f(z)-f(z_0)}{z-z_0}\cdot(z-z_0)\right] = f'(z_0)\cdot 0 = 0.$$

## Constancy in a Region

The Cauchy–Riemann equations have the following theoretical consequence.

**Theorem 123.** *Let $f \in H(D(0;R))$. Then any of the following conditions imply that $f$ is a constant function on $D(0;R)$:*
*a) $f'(z) = 0$ in $D(0;R)$.*
*b) $|f(z)| = c$ in $D(0;R)$.*
*c) $f(z)$ is real for all $z \in D(0;R)$.*

*Proof.* To prove a) let $f(z) = u(x, y) + iv(x, y)$ and let $z = x + iy$ in $D(0; R)$. Then from the proof of the Cauchy–Riemann equations we know

$$f'(z) = u_x + iv_x = -iu_y + v_y = 0,$$

which implies $u_x = v_x = u_y = v_y = 0$. Now fix $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$ in $D(0; R)$, we want to show $f(z_1) = f(z_2)$ (Figure 3.14).

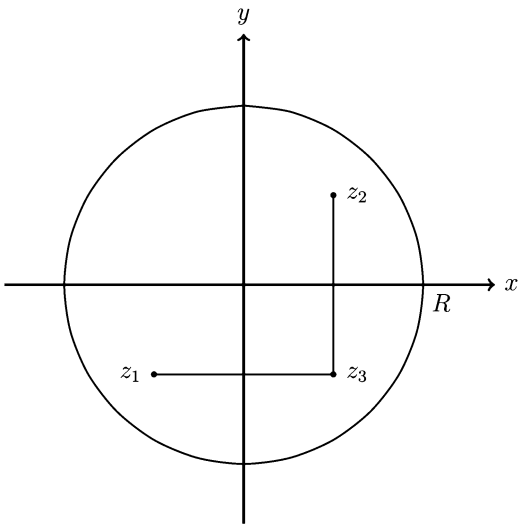



Figure 3.14: Constancy in a region

At least one of $z_3 = x_2 + iy_1$ or $z_4 = x_1 + iy_2$ lies in $D(0; R)$; without loss of generality suppose $z_3$ does. Then, each real-valued function $\phi : x \mapsto u(x, y_1)$ has derivative $\phi'(x) = u_x = 0$. Thus $\phi$ is constant, i.e., $u(x_1, y_1) = u(x_2, y_1)$. Similarly one can show $u(x_2, y_1) = u(x_2, y_2)$. Since $u$ and $v$ are both constant on any horizontal and vertical line segments in $D(0; R)$ we obtain $f(z_1) = f(z_2) = f(z_3)$.
To prove b), let $|f(z)| = u^2 + v^2 = c^2$. Taking partials with respect to $x$ will give

$$uu_x + vv_x = 0.$$

Now apply the Cauchy–Riemann equations to obtain

$$uu_y - vv_y = 0.$$

Similarly, taking partials with respect to $y$ and applying the Cauchy–Riemann equations will yield

$$uu_y + vu_x = 0.$$

Elimination of $u_y$ yields $(u^2 + v^2)u_x = 0$ which implies $u_x = 0$. Similarly, all the other partials are equal to zero.

The proof of c) is similar. If $f$ is real valued, then $v = 0$, so that $v_x = v_y = 0$. By the Cauchy–Riemann equations, $u_x = u_y = 0$ too. Hence $f$ must be constant. □

## Exercises

1. Verify the Cauchy–Riemann equations for the function

$$f(z) = z^2 + 3z + 2.$$

2. Suppose that $f$ is a holomorphic function on the disc $D = \{z : |z| < 1\}$ and that $\operatorname{Re} f(z) = 3$ for all $z \in D$. Show that $f$ is a constant on $D$.

3. Suppose that $f$ is holomorphic on $G = \{z : \operatorname{Re} z > 1\}$ and that

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0$$

on $G$. Show that there is a real constant $\alpha$ and a complex constant $\beta$ such that $f(z) = -i\alpha z + \beta$ on $G$.

4. Show that $f(z) = \overline{z}$ is not holomorphic at any point of $\mathbb{C}$.

5. Show, by changing variables, that the Cauchy–Riemann equations in terms of polar coordinates become

$$\frac{\partial u}{\partial r} = \frac{1}{r}\frac{\partial v}{\partial \theta}, \qquad \frac{\partial v}{\partial r} = -\frac{1}{r}\frac{\partial u}{\partial \theta}.$$

6. Let $f(z)$ be defined as:

$$f(z) = \frac{x^4 - y^4}{x^3 + y^3} + i\,\frac{x^4 + y^4}{x^3 + y^3} \qquad \text{with } f(0) = 0.$$

Show that $f(z)$ is not differentiable at the origin, despite the fact that $u$ and $v$ satisfy the Cauchy–Riemann equations.

7. Suppose $f(z)$ and $\overline{f(z)}$ are holomorphic functions on an open set $G$. Show that $f(z)$ is constant in $G$.

8. Show that if $f(z)$ is holomorphic in an open set $G$ and if one of the functions $u(x, y)$, $v(x, y)$ is constant, then the other one is also constant.

9. Let $G$ be an open subset of $\mathbb{C}$ and $G^* = \{z : \overline{z} \in G\}$. Suppose that $f$ is holomorphic on $G$, and define a function $g$ on $G^*$ by

$$g(z) = \overline{f(\overline{z})}.$$

Show that $g$ is holomorphic on $G^*$.

10. Where is $f(z) = \dfrac{z^3 + 2z + 1}{z^3 + 1}$ holomorphic? Compute the derivative.

11. Let $f = u + iv$ be holomorphic in an connected open set $D$, where $u$ and $v$ are real valued. Suppose that there are real constants $a, b$, and $c$ such that $a^2 + b^2 \neq 0$ and $au + bv = c$ in $D$. Prove that $f$ is constant in $D$.

12. Does there exists functions $f(z)$ and $g(z)$ that are holomorphic at $z = 0$ and that satisfy

1. $f\left(\frac{1}{n}\right) = f\left(\frac{-1}{n}\right) = \frac{1}{n^2}$ for $n = 1, 2, \ldots$
2. $g\left(\frac{1}{n}\right) = g\left(\frac{-1}{n}\right) = \frac{1}{n^3}$ for $n = 1, 2, \ldots$?

## 3.3 Power Series

### Convergence of Power Series

In general a *power series* is an expression of the form

$$\sum_{n=0}^{\infty} a_n z^n,$$

where $a_n \in \mathbb{C}$. Since absolute convergence implies convergence, to check convergence of a power series, we check the convergence of

$$\sum_{n=0}^{\infty} |a_n||z^n|.$$

Observe that if a given power series converges for some $z_0$, then it will also converge for all $z$ in the disc $|z| < |z_0|$.

In fact, there always exists an open disc on which a power series converges absolutely. The *radius of convergence* of a power series $\sum_{n=0}^{\infty} a_n z^n$ is defined as:

$$R := \sup\left\{|z| : \sum |a_n z^n| \text{ converges}\right\}.$$

If we use the convention that $1/0 = \infty$ and $1/\infty = 0$, then one can express $R$ as *Hadamard's formula* (see [37], pp. 55–56):

$$\frac{1}{R} = \limsup |a_n|^{1/n}.$$

The region $|z| < R$ is called the *disk of convergence.*

**Theorem 124.** *Let $\sum a_n z^n$ be a power series with radius of convergence $0 \leq R \leq \infty$. Then*

*(1) $\sum a_n z^n$ converges absolutely for all $z$, $|z| < R$.*
*(2) $\sum a_n z^n$ diverges for any $z$ with $|z| > R$.*

*On the boundary of the disc of convergence, $|z| = R$, a given power series either converges or diverges.*

*Proof.* To prove (1), if $|z| < R$, letting $\epsilon := R - |z| > 0$, observing $R - \epsilon$ is strictly less than $R$ and using the definition of supremum, there is some $z_1$ with

$$|z| = R - \epsilon < |z_1| \leq R$$

for which $\sum |a_n z_1|$ converges. Then

$$|a_n z^n| \leq |a_n z_1^n|$$

for all $n$ implies that $\sum |a_n z^n|$ converges by the comparison test.

The proof of (2) is done by contradiction. Suppose there is some $z$ with $|z| > R$ for which $\sum a_n z^n$ converges. Then the general term $|a_n z^n|$ is bounded, i.e., $\exists M$ such that $|a_n z^n| \leq M$ for all $n$. Now, pick $z_1$ such that $R < |z_1| < |z|$ so

$$|a_n z_1^n| = \left| a_n z^n \cdot \frac{z_1^n}{z^n} \right| \leq M \left| \frac{z_1}{z} \right|^n.$$

Since $|z_1/z| < 1$, the geometric series

$$\sum_{n=0}^{\infty} \left| \frac{z_1}{z} \right|^n$$

converges. Hence by the comparison test $\sum |a_n z_1^n|$ converges. This contradicts the definition of $R$. □

In the following examples we see that $R = 1$ for the geometric series while $R = \infty$ for the exponential function.

## Basic Examples

(1) The complex *exponential* function, for $z \in \mathbb{C}$, is defined as

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}.$$

When $z$ is real this definition coincides with the usual exponential function. Since

$$\left| \frac{z^n}{n!} \right| = \frac{|z|^n}{n!}$$

and

$$\sum_{n=0}^{\infty} \frac{|z|^n}{n!} = e^{|z|} < \infty,$$

the series above given for $e^z$ converges absolutely for all $z \in \mathbb{C}$.

It can be shown that $e^z$ is holomorphic for all $z \in \mathbb{C}$ (such functions are called *entire functions*) and its derivative is

$$\frac{d}{dz}(e^z) = \sum_{n=0}^{\infty} \frac{nz^{n-1}}{n!} = \sum_{m=0}^{\infty} \frac{z^m}{m!} = e^z.$$

We are allowed to differentiate term by term since the series defining $e^z$ is uniformly convergent in every disc in $\mathbb{C}$.

(2) The trigonometric functions $\sin z$ and $\cos z$ can also be given by complex series that converge in the whole complex plane. These are

$$\sin z = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n+1}}{(2n+1)!}$$

and

$$\cos z = \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n}}{(2n)!}.$$

Observe that the above expansions agree with the usual sine and cosine of the real argument. Using the fact that

$$e^{iz} = \cos z + i \sin z$$

and

$$e^{-iz} = \cos z - i \sin z,$$

one can obtain the so-called *Euler formulas* for $\sin z$ and $\cos z$ as follows:

$$\sin z = \frac{e^{iz} - e^{-iz}}{2i} \ , \ \cos z = \frac{e^{iz} + e^{-iz}}{2}.$$

(3) The *geometric series* $\sum_{n=0}^{\infty} z^n$ converges absolutely only on the disc $|z| < 1$, and for those points in $|z| < 1$,

$$\sum_{n=0}^{\infty} z^n = \frac{1}{1-z},$$

and $\dfrac{1}{1-z}$ is holomorphic in an open set $D \setminus \{1\}$.

Note that the expansion

$$\frac{1}{1-z} = 1 + z + z^2 + \cdots + z^n + \cdots \qquad \text{for } |z| < 1$$

can be used to derive expansion for other series. For example:

$$\frac{1}{1+z} = \frac{1}{1-(-z)} = 1 + (-z) + (-z)^2 + \cdots + (-z)^n + \cdots = \sum_{n=0}^{\infty} (-1)^n z^n \text{ for } (|z| < 1),$$

$$\frac{1}{1-z^2} = 1 + z^2 + (z^2)^2 + \cdots + (z^2)^n + \cdots = \sum_{n=0}^{\infty} (z^2)^n \text{ for } (|z| < 1),$$

$$\frac{1}{z-2} = \frac{1}{2\left(\frac{z}{2} - 1\right)} = -\frac{1}{2} \cdot \frac{1}{1 - \frac{z}{2}} = -\frac{1}{2} \sum_{n=0}^{\infty} \left(-\frac{z}{2}\right)^n = \sum_{n=0}^{\infty} (-1)^{n+1} \frac{z^n}{2^{n+1}}$$

(for $|z/2| < 1$ or $|z| < 2$).

## Differentiating Power Series

There is a vital connection between power series and holomorphic functions, as we illustrate below. To establish $\sum a_n z^n$ is differentiable for $|z| < R$, we observe that even though there is an obvious candidate for the derivative $\sum n a_n z^{n-1}$ (supposing we can differentiate term by term), it is not even clear that $\sum n a_n z^{n-1}$ converges for $|z| < R$. We first prove this in the following lemma.

**Lemma 15.** *The power series $\sum a_n z^n$ and $\sum n a_n z^{n-1}$ have the same radius of convergence.*

*Proof.* Suppose $\sum |a_n z^n|$ converges for $|z| < R$. Choose $r$ such that $|z| < r < R$ and assume $z \neq 0$. Then

$$|n a_n z^{n-1}| = \frac{n}{|z|} \frac{|z|^n}{r^n} |a_n| r^n.$$

Since $|z|/r < 1$, the series

$$\sum_{n=0}^{\infty} n \left(\frac{|z|}{r}\right)^n$$

converges by the ratio test, thus

$$\exists M \quad \text{such that} \quad n\left(\frac{|z|}{r}\right)^n \leq M.$$

Hence,

$$|na_n z^{n-1}| \leq \frac{M}{|z|}|a_n r^n|,$$

and now applying the comparison test yields the claim. To show the converse, assume $\sum na_n z^{n-1}$ converges. Then from

$$|a_n z^n| \leq |z||na_n z^{n-1}| \qquad (n \geq 1)$$

the result follows by the comparison test. □

The following theorem provides a very important class of holomorphic functions that are quite easy to work with.

**Theorem 125.** *The power series*

$$f(z) = \sum_{n=0}^{\infty} a_n z^n$$

*defines a holomorphic function in its disc of convergence. The derivative $f'$ of $f$ is also a power series obtained by differentiating term by term the series for $f$. Namely*

$$f'(z) = \sum_{n=0}^{\infty} na_n z^{n-1}.$$

*Furthermore, $f'(z)$ has the same radius of convergence as $f(z)$.*

*Proof.* Let $f(z) = \sum_{n=1}^{\infty} a_n z^n$ have radius of convergence $R > 0$, then $f$ is holomorphic on $D(0;R)$. By Lemma 15 we know the power series $\sum a_n z^n$ and $\sum na_n z^{n-1}$ have the same radius of convergence. Thus we can define

$$g(z) := \sum_{n=1}^{\infty} na_n z^{n-1} \quad \text{for} \quad |z| < R$$

and we want to show that $f'(z)$ exists and equals $g(z)$ for $z \in D(0;R)$. With this in mind for $z$, $z_0 \in D(0;R)$ form

$$\frac{f(z) - f(z_0)}{z - z_0} - g(z_0) = \sum_{n=1}^{\infty} a_n \left(\frac{z^n - z_0^n}{z - z_0} - nz_0^{n-1}\right),$$

and show that this tends to 0 as $z \to z_0$. We now observe

$$\frac{z^n - z_0^n}{z - z_0} = \frac{(z - z_0)(z^{n-1} + z_0 z^{n-2} + \cdots + z_0^{n-1})}{z - z_0}$$

and therefore

$$\begin{aligned}\frac{f(z) - f(z_0)}{z - z_0} - g(z_0) &= \sum_{n=1}^{\infty} a_n \left(z^{n-1} + z_0 z^{n-2} + \cdots + z_0^{n-1} - n z_0^{n-1}\right) \\ &= \sum_{I} + \sum_{II},\end{aligned}$$

here by $\sum_I$ we mean $\sum_{n=1}^{N} a_n \left(z^{n-1} + z_0 z^{n-2} + \cdots + z_0^{n-1} - n z_0^{n-1}\right)$ and $\sum_{II}$ denotes the rest of the series, i.e., $\sum_{n=N+1}^{\infty} a_n \left(z^{n-1} + z_0 z^{n-2} + \cdots + z_0^{n-1} - n z_0^{n-1}\right)$.

Given $\epsilon > 0$ choose $r$ such that $|z_0| < r < R$. Then $\sum_{n=1}^{\infty} n a_n r^{n-1}$ is convergent so by the property of convergent series its tail can be controlled and if we choose $z$ close enough to $z_0$ to ensure that $|z| < r$ there must exist $N(\epsilon)$ such that $\left|\sum_{II}\right| < \frac{\epsilon}{2}$. Furthermore note that when $z \to z_0$ the term $\left(z^{n-1} + z_0 z^{n-2} + \cdots + z_0^{n-1}\right)$ is $n z_0^{n-1}$ and $\sum_I$ is a polynomial in $z$ so $\sum_I \to 0$ as $z \to z_0$ and thus $\left|\sum_I\right| < \frac{\epsilon}{2}$. We now ensure that $z$ is close enough to $z_0$ so that

$$\left|\frac{f(z) - f(z_0)}{z - z_0} - g(z_0)\right| = \left|\sum_I + \sum_{II}\right| \le \left|\sum_I\right| + \left|\sum_{II}\right| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

and so $f'(z_0) = g(z_0)$ as claimed. For more technical details of the proof one can consult ([50], pp. 16–17). □

One of the important consequences of the above result comes from successive applications of this theorem, which implies that a power series is infinitely differentiable in its disc of convergence, i.e., if $f(z) = \sum a_n z^n$ for $|z| < R$, then

$$f'(z) = \sum_{n=1}^{\infty} n a_n z^{n-1} \text{ for } |z| < R$$

$$f''(z) = \sum_{n=2}^{\infty} n(n-1) a_n z^{n-2} \text{ for } |z| < R$$

and so on. Note $f$ has derivatives of all orders in $|z| < R$.

**Corollary 26.** *All the higher derivatives $f', f'', \ldots, f^{(n)}, \ldots$ of the power series $f(z) = \sum a_n z^n$ exist for all $z$ within the disc of convergence and*

$$f^{(k)}(0) = k! a_k.$$

*Proof.* To see the expression for $f^{(k)}(0)$, note that if

$$f(z) = \sum_{n=0}^{\infty} a_n z^n = a_0 + a_1 z + a_2 z^2 + a_3 z^3 + \cdots + a_n z^n + \cdots$$

then

$$f'(z) = a_1 + 2a_2 z + 3a_3 z^2 + \cdots + n a_n z^{n-1} + \cdots$$
$$f''(z) = 2a_2 + 2 \cdot 3 a_3 z^2 + \cdots + n(n-1) a_n z^{n-2} + \cdots$$

and thus

$$f(0) = a_0,\ f'(0) = a_1,\ f''(0) = 2a_2, \ldots, f^{(k)}(0) = k! a_k$$

or equivalently

$$a_k = \frac{f^{(k)}(0)}{k!}.$$

□

**Remark.** This theorem highlights the basic difference between real and complex functions. As you remember, in the real case, one encounters functions which are differentiable $n$ times, but not $(n+1)$ times. For example, if we take

$$f(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ x^2 & \text{if } x > 0 \end{cases},$$

one can see that $f'(0)$ exists but $f''(0)$ does not exist. In the complex case, if the complex derivative $f'$ of $f$ exists, then all higher derivatives exist.

For another distinction, consider the function $f$ defined on $\mathbb{R}$ by:

$$f(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ e^{-1/x^2} & \text{for } x > 0 \end{cases}.$$

One can prove that $f$ is infinitely differentiable on $\mathbb{R}$ and $f^{(n)}(0) = 0$. Thus the Taylor series expansion of this function at $x = 0$ is

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n = 0 \neq f(x).$$

In the complex case, if a function is differentiable in the sense of complex variables then the validity of the Taylor expansion of that function is guaranteed. In this sense, complex analysis is "simple" relative to the real case.

## Exercises

1. Suppose that the complex series $\sum_{n=1}^{\infty} a_n$ converges. Show that $\lim_{n\to\infty} a_n = 0$.

2. Determine whether the given infinite series converges or diverges:

   a) $\sum_{n=1}^{\infty} \left(\frac{1+2i}{\sqrt{6}}\right)^n$

   b) $\sum_{n=1}^{\infty} \frac{i^n}{n}$

3. Show that each of the following series converges for all $z$:

   a) $\sum_{n=0}^{\infty} \frac{z^n}{n!}$ b) $\sum_{n=0}^{\infty} (-1)^n \frac{z^{2n}}{(2n)!}$ c) $\sum_{n=0}^{\infty} (-1)^n \frac{z^{2n+1}}{(2n+1)!}$

4. Prove that if $a_n$ for $n = 0, 1, \ldots$ are complex numbers such that $\sum |a_n|$ converges, then

$$\left| \sum_{n=0}^{\infty} a_n \right| \leq \sum_{n=0}^{\infty} |a_n|.$$

5. By differentiating an appropriate series, show that

$$\sum_{k=1}^{\infty} k z^k = \frac{z}{(1-z)^2} \quad \text{for} \quad |z| < 1.$$

6. Write down an expansion of the form $\sum_{n=0}^{\infty} a_n z^n$ for

   a) $\frac{1}{3z+4}$

   b) $\frac{1+iz}{1-iz}$

7. Show that if $\{a_n\}$ is a sequence of nonzero complex numbers such that

$$\lim_{n\to\infty} \frac{|a_{n+1}|}{|a_n|} = L,$$

then

$$\lim_{n\to\infty} |a_n|^{\frac{1}{n}} = L.$$

8. Find a series expansion of $(1-z)^{-1}$ valid in $|z+3| < 4$.

9. Show that for $a \neq b$,

$$\frac{1}{(z-a)(z-b)} = \frac{1}{(a-b)}\left(\sum_{n=0}^{\infty}\left(\frac{1}{b^{n+1}} - \frac{1}{a^{n+1}}\right)z^n\right)$$

is true for $|z| < \min\{|a|, |b|\}$.

10. Prove the following:

a) The power series $\sum \frac{z^n}{n^2}$ converges at every point of the unit circle.

b) The power series $\sum nz^n$ does not converge on any point of the unit circle.

c) The power series $\sum \frac{z^n}{n}$ converges at every point of the unit circle except at $z = 1$.

11. Determine all holomorphic functions $f$ defined on the unit disc $D$ which satisfy

$$f''\left(\frac{1}{n}\right) + f\left(\frac{1}{n}\right) = 0.$$

12. Let the sequence $a_0, a_1, \ldots$ be defined such that

$$1 - x^2 + x^4 - x^6 + \cdots = \sum_{n=0}^{\infty} a_n(x-3)^n \quad 0 < x < 1.$$

Find $\text{limsup}_{n\to\infty}\left(|a_n|^{\frac{1}{n}}\right)$.

13. Suppose $f(z) = \sum_{n=0}^{\infty} a_n z^n$ has a radius of convergence $R > 0$. Show that $g(z) = \sum_{n=0}^{\infty} \frac{a_n z^n}{n!}$ is a holomorphic function on $\mathbb{C}$ and that for $0 < r < R$, there is a constant $M$ such that

$$|g(z)| \leq Me^{\frac{|z|}{r}}.$$

## 3.4 Some Holomorphic Functions

### The Exponential Function

In the section on power series we already showed that the power series $\sum \frac{z^n}{n!}$ has an infinite radius of convergence and we may define the *exponential function*

as:

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!} \quad \text{for } z \in \mathbb{C},$$

so $e^z$ is holomorphic and hence also continuous in $\mathbb{C}$. For $z = x + iy$,

$$e^z = e^{x+iy} = e^x(\cos y + i \sin y),$$

and we see that

$$f(z) = e^z = u(x, y) + iv(x, y),$$

where

$$u(x, y) = e^x \cos y \text{ and } v(x, y) = e^x \sin y.$$

The functions $u(x, y)$ and $v(x, y)$ are continuous and real valued with continuous first-order partial derivatives and satisfy the Cauchy–Riemann equations,

$$u_x = e^x \cos y = v_y \text{ and } u_y = -e^x \sin y = -v_x,$$

$$\frac{d}{dz}(e^z) = \frac{\partial u}{\partial x} + i\frac{\partial v}{\partial x} = e^x \cos y + ie^x \sin y = e^z.$$

**Modulus, Argument, and Conjugate of $e^z$:**

$$\text{If } w = e^z = e^x e^{iy} \text{ then } |w| = |e^z| = e^x \text{ and } \arg(e^z) = y + 2k\pi.$$

Since $e^x > 0$ for all $x$, we know $|e^z| > 0$. In fact, $e^z \neq 0$ for all $z$. The value $w = 0$ is not in the range of $e^z$, however as we know $e^{\pi i} = -1$, $e^z$ could be negative. For the conjugate, it is easy to see that

$$\overline{e^z} = e^{\overline{z}}.$$

The following are some algebraic properties of $e^z$. Let $z_1$ and $z_2$ be complex numbers, then

a) $e^0 = 1$.

b) $e^{z_1} \cdot e^{z_2} = e^{z_1+z_2}$.

c) $\dfrac{e^{z_1}}{e^{z_2}} = e^{z_1-z_2}$.

d) $(e^{z_1})^n = e^{nz_1} \qquad n = 0, \pm 1, \pm 2, \ldots$.

e) $e^z$ is a periodic function with period $2\pi i$.

To show e) above, consider $e^{z+2\pi i} = e^z \cdot e^{2\pi i} = e^z$. Repeating this process, one can say $e^{z+4\pi i} = e^z$ and thus $e^{z+2n\pi i} = e^z$ for $n = 0, \pm 1, \pm 2, \ldots$ which in turn asserts that $e^z$ is not a one-to-one function. In other words,

$$e^{z_1} = e^{z_2} \iff z_1 = z_2 + 2n\pi i \ \ (n \in \mathbb{Z}).$$

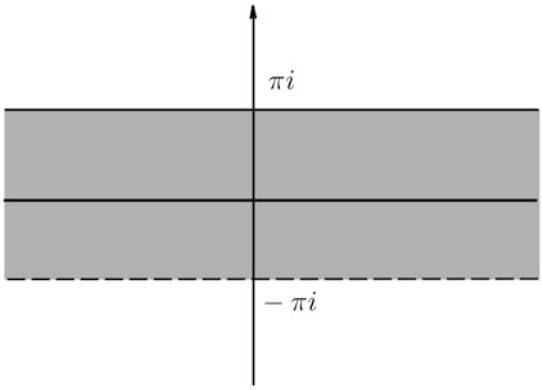



Figure 3.15: Fundamental region

To find the inverse function of $e^z$, first we divide the complex plane into horizontal strips by taking

$$-\infty < x < \infty \text{ and } -\pi < y \leq \pi.$$

This infinite strip is called the *fundamental region* (Figure 3.15) and if we suppose all values of $e^z$ are assumed in the fundamental region then $e^z$ is one-to-one.
Furthermore, if we look at a vertical line segment, say $z = a + it$ , $-\pi < t \leq \pi$, then the image of this vertical line

$$w = e^z = e^{a+it} = e^a \cdot e^{it}$$

is a circle centered at 0 radius $e^a$ (Figure 3.16).

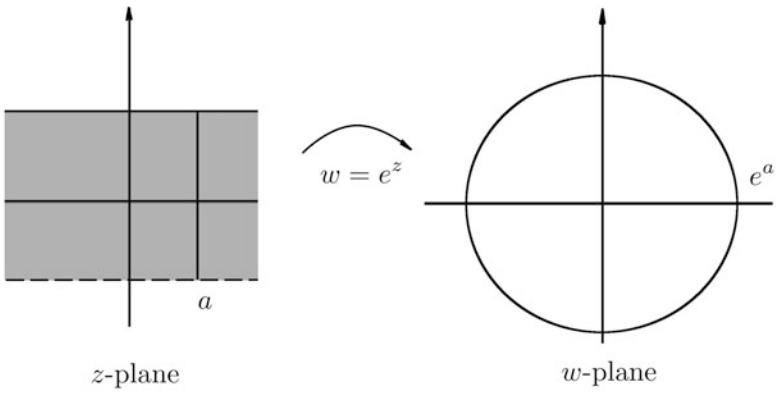



Figure 3.16: Image of a vertical line segment under $w = e^z$

Images of the vertical line segments in $(-\pi, \pi]$ are circles.
On the other hand if we take a horizontal line in the fundamental region, the image of this line under $w = e^z$ is a ray (Figure 3.17).

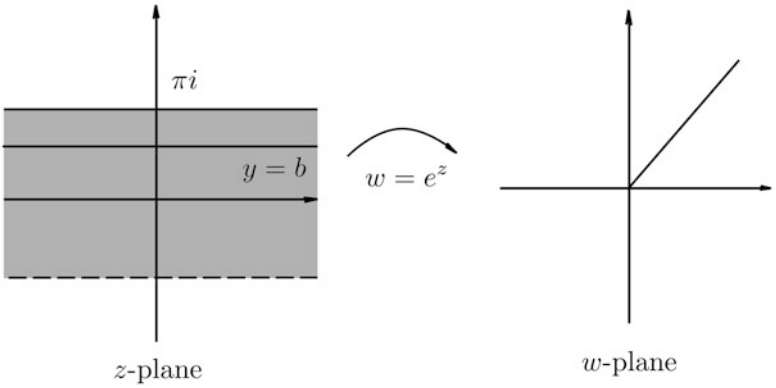



Figure 3.17: Image of a horizontal line under $w = e^z$

Since the only $z$ on the horizontal line $y = b$ is of the form $z = x + ib$ where $-\infty < x < \infty$,

$$w = e^z = e^{x+ib} = e^x \cdot e^{ib}.$$

We define a new parameter $e^x = s$ and observe that $0 < s < \infty$ since $-\infty < x < \infty$. Using the parameter $s$, the image is given by

$$w = se^{ib} \quad \text{where} \quad 0 < s < \infty,$$

which is the set consisting of all points $w \neq 0$ in the ray starting from 0 containing the point $e^{ib} = \cos b + i \sin b$.

## Complex Logarithmic Function

The polar representation of a complex number $z = |z|e^{i\theta}$ is not uniquely determined. This is the main reason for multi-valuedness in complex analysis. For $z \neq 0$, we define the argument of $z$ to be the set of values $\{\theta \in \mathbb{R} : z = |z|e^{i\theta}\}$. In fact, when we say argument of $z$, we mean not a single number, but an infinite set. More precisely we can define $\operatorname{Arg} z$ (note the capital letter) as the set of values

$$\operatorname{Arg} z = \{\theta + 2k\pi \quad \text{for} \quad k \in \mathbb{Z}\}.$$

For example,

$$\begin{aligned}\operatorname{Arg} i &= \frac{\pi}{2} + 2k\pi \\ &= (4k+1)\frac{\pi}{2}\ ,\ k \in \mathbb{Z}.\end{aligned}$$

If we restrict $\theta$ to $(-\pi, \pi]$, we can have a single value for $\operatorname{Arg} z$; we call this the *principal value* of the argument and use the notation $\arg\, z$. It should be emphasized that the choice of the interval $(-\pi, \pi]$ is completely arbitrary: we might have chosen the interval $[0, 2\pi)$ instead.

In real analysis, the statements $y = e^x$ and $x = \log y$ where $y > 0$ are equivalent. If we try to use this approach to define $\log\, z$ for a complex $z$, we encounter a difficulty. To define complex logarithms as an inverse of an exponential requires some care, since $e^w = z$ has infinitely many solutions. Let us suppose that $w = \log z$ is equivalent to $z = e^w$, where $w = u + iv$. Then

$$z = e^w = e^{u+iv} = e^u \cdot e^{iv}$$

and so

a) $e^u = |z|$. Thus $u = \log |z|$ (we are using a simple log to denote $\log_e$).

b) $v =$ argument of $z$ .

Thus $e^w = z$ will yield $w = \log |z| + i\operatorname{Arg} z$. However we just noted that there are infinitely many Arg$z$, thus when we write

$$\operatorname{Log} z = \{\log |z| + i\theta :\ \theta \in \operatorname{Arg}(z)\}$$

we again mean a multi-valued function and restricting $\theta$ to either $[0, 2\pi)$ or $(-\pi, \pi]$ gives a single value which we call the *principle value or principal branch of the logarithmic function* expressed as

$$\log z = \log |z| + i \arg z.$$

It follows that the choice of principal logarithm is arbitrary; however, once we make the choice of $(-\pi,\ \pi]$ , we shall stick to it. When we restrict $\theta$ to $(-\pi, \pi]$, we basically are crossing out the negative $x$-axis. We think of the plane as having a cut along the negative $x$-axis, preventing the argument of $z$ from leaving the interval $(-\pi,\ \pi]$ (Figure 3.18).

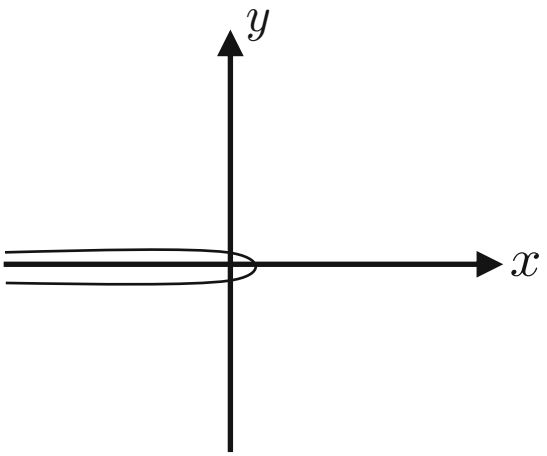



Figure 3.18: $\mathbb{C} \setminus (-\infty, 0]$: branch points and the cut

In the definition of the principal argument and the principal logarithm, the position of the cut is arbitrary. The positive $x$-axis would do as well so would any half line containing the origin. The key issue here is that the cut should contain 0 and should go off the infinity. We say 0 and $\infty$ are *branch points*. If $z$ moves around any closed path contained in $\mathbb{C} \setminus (-\infty, 0]$, the logarithm changes continuously and returns to its original value.

**Proposition 41.** *Let* $D^* = \mathbb{C} \setminus (-\infty, 0]$. *Then*

*a)* $\log z$ *is continuous on* $D^*$.

*b)* $\log z$ *is holomorphic on* $D^*$ *and if* $f_1(z) =$ *principal value of* $\log z$, *then* $f_1'(z) = \dfrac{1}{z}$.

*Proof.* a) Notice that when we take the principal value of

$$\begin{aligned} \log z &= u(x,y) + iv(x,y) \\ &= \log|z| + i \arg z, \end{aligned}$$

$u(x,y) = \log|z| = \log\sqrt{x^2+y^2}$ is continuous everywhere except at $(0,0)$, and $v(x,y) = \arg z$ is discontinuous at every point on the negative $x$-axis.

b) Let $z = re^{i\theta}$ $-\pi < \theta \leq \pi$. Then

$$\log z = \log r + i\theta = u(r,\theta) + iv(r,\theta),$$

where $u(r,\theta) = \log r$ and $v(r,\theta) = \theta$. Taking partial derivatives, we have

$$\frac{\partial u}{\partial r} = \frac{1}{r}, \qquad \frac{\partial u}{\partial \theta} = 0, \qquad \frac{\partial v}{\partial r} = 0 \quad \text{and} \quad \frac{\partial v}{\partial \theta} = 1.$$

Recalling that the Cauchy–Riemann equations expressed in polar coordinates take the form:

$$\frac{\partial u}{\partial r} = \frac{1}{r}\frac{\partial v}{\partial \theta} \quad \text{and} \quad \frac{\partial v}{\partial r} = -\frac{1}{r}\frac{\partial u}{\partial \theta},$$

one can easily check that the Cauchy–Riemann equations are satisfied. Moreover,

$$\frac{\partial u}{\partial r}, \quad \frac{\partial u}{\partial \theta}, \quad \frac{\partial v}{\partial \theta}, \quad \frac{\partial v}{\partial r}$$

are all continuous in $D^*$. Thus, $\log z$ is holomorphic in $D^*$. Again, denoting the principal value by $f_1(z) = \log z$,

$$f_1'(z) = e^{-i\theta}\left(\frac{\partial u}{\partial r} + i\frac{\partial v}{\partial r}\right) = e^{-i\theta}\left(\frac{1}{r} + i \cdot 0\right) = \frac{1}{re^{i\theta}} = \frac{1}{z}.$$

Thus, if $|z| > 0$ and $-\pi < \arg z \leq \pi$, then

$$f_1'(z) = \frac{1}{z}. \qquad \square$$

**Remark.** The above discussion shows that $\log z$ is the inverse of $e^z$ in the following sense:

a) For any branch of $\log z$ we have $e^{\log z} = z$, $z \neq 0$.

b) If we choose a branch lying in $[0, 2\pi)$ or $(-\pi, \pi]$ then $\log e^z = z$, for $z = x + iy$ where $-\infty < x < \infty$ and $-\pi < y \leq \pi$.

**Powers.**
Another important multi-function is $z^{1/n}$, where $n$ is a positive integer. We have seen if $z \neq 0$ and $n$ is a positive integer then $w^n = z$ has $n$ solutions. If $z = re^{i\theta}$ as a multi-function,

$$w = z^{1/n} = \left\{ r^{1/n} e^{\frac{(\theta + 2k\pi)i}{n}} : \quad k = 0, 1, 2, \ldots, n-1 \right\}.$$

If $\arg z$ is the principal value then the principal value of the $n$th root is $z^{1/n} = r^{1/n} e^{i \arg z/n}$. Suppose $\alpha \in \mathbb{C}$ and $z \neq 0$; now we want to determine $z^\alpha$. Since $z^\alpha = e^{\alpha \log z}$, we can write a set of values for $z^\alpha$, namely

$$z^\alpha = e^{\alpha[\log|z| + i \arg z]}.$$

Again one can determine the principal value of $z^\alpha$ by restricting $\theta$. For example, we would like to find $(-i)^i$. Assume this time $\arg z \in (0, 2\pi]$, then,

$$\begin{aligned}(-i)^i &= e^{i \log(-i)} \\ &= e^{i(\log|i| + i\frac{3\pi}{2})} \\ &= e^{i(\log 1 + i\frac{3\pi}{2})} \\ &= e^{-3\pi/2}.\end{aligned}$$

**Example 100** (Riemann surface for $y = e^z$)**.** The theory of Riemann surfaces provide a framework for a study of multi-functions such as $w = e^z$, $\log z$ or $\sqrt{z}$. The main idea is instead of working with individual branches of a multi-function in a cut plane, consider all branches as a single function on a domain set consisting of many copies of the plane. Since we will glue these copies together, one can pass continuously from one branch of the multi-function to another. To clarify, let us consider a Riemann surface for $w = e^z$ on the plane $x \leq 0$. Define the fundamental region as an infinite strip $S_n$ where

$$S_n = \{(x, y) : \ x \leq 0, \quad (2n-1)\pi < y \leq (2n+1)\pi\}$$

for $n = 0, \pm 1, \pm 2, \ldots$ Notice that $w = e^z$ maps each infinite strip $S_n$ onto the punctured unit disk $D_n = 0 < |w| \leq 1$ shown in Figure 3.19. In other words, $e^z$

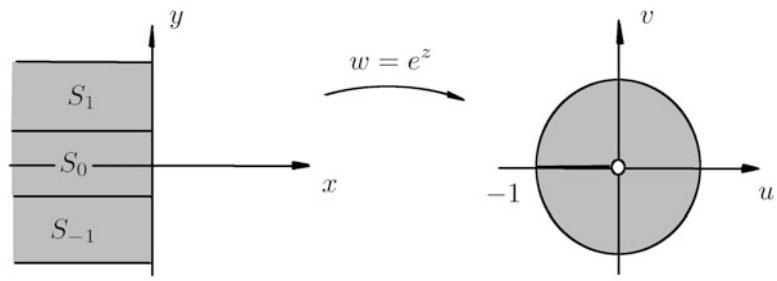



Figure 3.19: $z \mapsto e^z$

describes an infinite-to-one covering of the punctured disk $D_n$. To visualize this covering, imagine that there is a different image disk $D_n$ for each half infinite strip $S_n$. Now cut each disk $D_n$ open along $-1 \leq u < 0$. To form the Riemann surface for $w = e^z$ attach for each $n$, the cut disk $D_n$ to the cut disk $D_{n+1}$ along the edge. Think of this surface in $xyz$-space so that for each $z$ in the half-plane, the images $\ldots z_{-1}, z_0, z_1, \ldots$ of $z$ in $\ldots D_{-1}, D_0, D_1, \ldots$, respectively lie directly above the point $w = e^z$ in the $xy$-plane. In other words if one projects the points on the Riemann surface vertically down onto the $xy$-plane one can see the infinite-to-one nature of the mapping $w = e^z$ (Figure 3.20). Readers interested in pursuing the study of Riemann surfaces will have no difficulty finding relevant literature, e.g., [47].

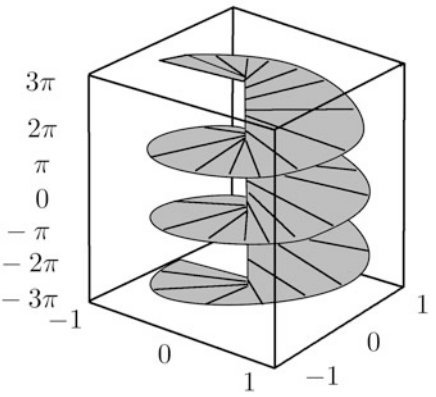



Figure 3.20: Riemann surface for $w = e^z$

## Trigonometric and Hyperbolic Functions

For $z \in \mathbb{C}$, we define:

$$\cos z := \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n}}{(2n)!} \ ; \ \sin z := \sum_{n=0}^{\infty} (-1)^n \frac{z^{2n+1}}{(2n+1)!} \quad \text{and}$$

$$\cosh z := \sum_{n=0}^{\infty} \frac{z^{2n}}{(2n)!} \ ; \ \sinh z := \sum_{n=0}^{\infty} \frac{z^{2n+1}}{(2n+1)!}.$$

Since all of these four functions are holomorphic in $\mathbb{C}$, we can find their derivatives differentiating term by term, and it can be shown that

$$\frac{d}{dz}(\cos z) = -\sin z \ , \ \frac{d}{dz}(\sin z) = \cos z,$$

$$\frac{d}{dz}(\cosh z) = \sinh z \ , \ \frac{d}{dz}(\sinh z) = \cosh z.$$

From $e^{iz} = \cos z + i \sin z$, we also have

$$\cos z = \frac{e^{iz} + e^{-iz}}{2} \text{ and } \sin z = \frac{e^{iz} - e^{-iz}}{2i},$$

similarly

$$\cosh z = \frac{e^{z} + e^{-z}}{2} \text{ and } \sinh z = \frac{e^{z} - e^{-z}}{2}.$$

**Proposition 42.** *We have the following equalities:*

*a)* $\sin^2 z + \cos^2 z = 1$.

*b)* $\sin(z+w) = \sin z\, \cos w + \cos z\, \sin w$.

*c)* $\cos(z+w) = \cos z\, \cos w - \sin z\, \sin w$.

*Proof.* This follows directly from the definitions. For example

$$\begin{aligned}\sin^2 z + \cos^2 z &= \left(\frac{e^{iz} - e^{-iz}}{2i}\right)^2 + \left(\frac{e^{iz} + e^{-iz}}{2}\right)^2 \\ &= \frac{e^{2iz} - 2 + e^{-2iz}}{4} + \frac{e^{2iz} + 2 + e^{-2iz}}{4} = 1.\end{aligned}$$

We leave parts b) and c) to the reader to verify. □

It is also quite straightforward to see that the complex trigonometric functions are periodic. Since if $f$ is holomorphic on some open set $G$, then $1/f$ is holomorphic in $G$ provided $f(z) \neq 0$ in $G$. Thus, it is good to locate zeros of the trigonometric functions. It turns out that

$$\cos z = 0 \iff z = \frac{1}{2}(2k+1)\pi \ , \ k \in \mathbb{Z}$$

$$\sin z = 0 \iff z = k\pi \ , \ k \in \mathbb{Z}$$

and

$$\cosh z = 0 \iff z = \frac{1}{2}(2k+1)\pi i \ , \ k \in \mathbb{Z}$$

$$\sinh z = 0 \iff z = k\pi i \ , \ k \in \mathbb{Z}.$$

**Remark** For $x \in \mathbb{R}$, $|\sin x| \leq 1$ and $|\cos x| \leq 1$ are quite useful inequalities. For $z \in \mathbb{C}$ these inequalities are no longer valid; in fact both functions $\sin z$ and $\cos z$ are unbounded in $\mathbb{C}$. See Exercise 4 at the end of this section.

## Exercises

1.

a) Find the real and imaginary parts of $e^{e^z}$.

b) Simplify $e^{z^2}$, $e^{iz}$, $e^{\frac{1}{z}}, z \neq 0$ where $z = x + iy$.

2. Suppose $f$ is holomorphic in $D(0\,;1)$. Prove that if $e^f$ is constant, then $f$ is constant.

3. Find the image of the given set under the exponential mapping:

a) The line $x = 3$.

b) The infinite strip $1 < x \leq 2$.

4. Establish the following identities:

   a) $\sin(z+w) = \sin z \cos w + \cos z \sin w$ for $z, w \in \mathbb{C}$.

   b) $\cos(x+iy) = \cos x \cosh y - i \sin x \sinh y$.

   c) $\sin(-z) = -\sin z, \quad \cos(-z) = \cos z, \quad \sin\left(\frac{\pi}{2} - z\right) = \cos z$.

   d) $\sin \overline{z} = \overline{\sin z}, \quad e^{\overline{z}} = \overline{e^z}$.

5. Use the equation $\sin z = \sin x \cosh y + i \sinh y \cos x$, where $z = x + iy$, to prove that

$$|\sinh y| \leq |\sin z| \leq |\cosh y|.$$

6. Find all solutions to the following equations :

   a) $\cosh z = -1$.

   b) $\tan z = i$.

7. Find all values of:

   a) $\log 1 \quad \text{and} \quad \log i$.

   b) $\log z$, where $z = \dfrac{5 + 5\sqrt{3}i}{2}$.

8. Compute the following values when the log is defined by its principal value on the open set $S$ equal to the plane with the positive real axis deleted.

   a) $\log(-1+i)$.

   b) $(-1)^i$.

9. Solve $\sin z = 5$.

10. Show that for $z_1, z_2 \in \mathbb{C} \setminus \{0\}$, the identity for the principal value of the logarithm

$$\log(z_1 z_2) = \log z_1 + \log z_2$$

holds true up to the integral multiples of $2\pi i$.

11. Find a domain on which $\log(3z - i)$ is holomorphic.

12.

a) Verify that $(z^{\alpha})^{n} = z^{n\alpha}$ for $z \neq 0$ and $n$ an integer.

b) Give an example to show that for $z \neq 0$ , we can have $(z^{\alpha_1})^{\alpha_2} \neq z^{\alpha_1\alpha_2}$.

## 3.5 Conformal Mappings

Suppose that $w = f(z)$ is a complex mapping defined on an open set $G$. We say the mapping $f$ is *conformal* at the point $z_0$ if it "preserves the angle" between any two curves intersecting at $z_0$. As we shall see later, every holomorphic function whose derivative is nonzero defines such mappings. These mappings, apart from their geometric interest, are used in advanced complex analysis, as well as solving certain physical problems such as problems in two-dimensional fluid flow.( See, e.g., [54], p. 263 and p. 386). The main motivation behind conformal mappings is to transform a given region $G$ onto a geometrically simpler region $G^*$, say the unit disc $D$ or upper half-plane $\mathbb{H}$, where many ideas have been developed to study holomorphic functions. In the following we will consider mapping a region whose boundary consists of lines, rays, circular arcs, and line segments. We will see that for many such cases, all we need is the following three basic types of maps and their certain linear combinations, namely:

a) Linear fractional transformations (also known as Möbius transformations)

b) Exponentials

c) Integer and non-integer powers

Before we understand angle preserving maps, we need to clarify what we mean by the angle between two curves $\gamma_1$ and $\gamma_2$.

### Angle Between Curves

Assume that $\gamma_1$ and $\gamma_2$ are smooth curves in $G$ that intersect at $z_0$ with the parametrization

$$\gamma_1(t) = (x_1(t), y_1(t)) \quad \text{and} \quad \gamma_2(t) = (x_2(t), y_2(t)).$$

Suppose the orientation of both curves are in the direction of increasing $t$, and since $z_0 = \gamma_1(t_0) = \gamma_2(t_0)$, the angle between these two curves is defined to be the angle between the tangent vectors to the curves (Figure 3.21).

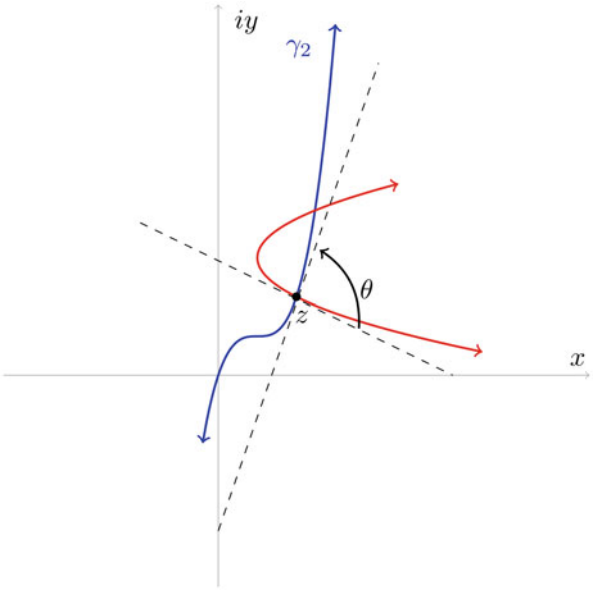



Figure 3.21: Angle between curves

Note that tangent vectors of the curves are

$$\frac{d\gamma_1}{dt} = \left(\frac{dx_1}{dt}, \frac{dy_1}{dt}\right) \quad \text{and} \quad \frac{d\gamma_2}{dt} = \left(\frac{dx_2}{dt}, \frac{dy_2}{dt}\right),$$

and we are interested in the dot product between the tangent vectors $\frac{d\gamma_1}{dt}$ and $\frac{d\gamma_2}{dt}$. Also the angle between $\gamma_1$ and $\gamma_2$ satisfies

$$\theta = \arg \gamma_1'(z_0) - \arg \gamma_2'(z_0)$$

in the interval $[0, \pi]$.

**Remark 69.** In general, the cosine of the angle between $z = \gamma_1$ and $w = \gamma_2$ can be found by considering the dot product of these two vectors. For given two complex numbers $z, w$ define the dot product by:

$$\langle z, w \rangle = \mathrm{Re}(z\overline{w}). \tag{3.1}$$

Then if $\theta$ is the angle between $z$ and $w$,

$$\cos\theta = \frac{\langle z, w\rangle}{|z||w|} \quad \text{and} \quad \sin\theta = \frac{\langle z, -iw\rangle}{|z||w|}$$

determines the angle.

**Definition 105.** Let $w = f(z)$ be a complex function defined in a domain $G$ and let $z_0$ be a point in $G$. Then we say $w = f(z)$ is *conformal at* $z_0$ if for every pair of smooth curves $\gamma_1$ and $\gamma_2$ intersecting at $z_0$ the angle between $\gamma_1$ and $\gamma_2$ is equal to the angle between the image curves $f(\gamma_1)$ and $f(\gamma_2)$ at $f(z_0)$ (Figure 3.22).

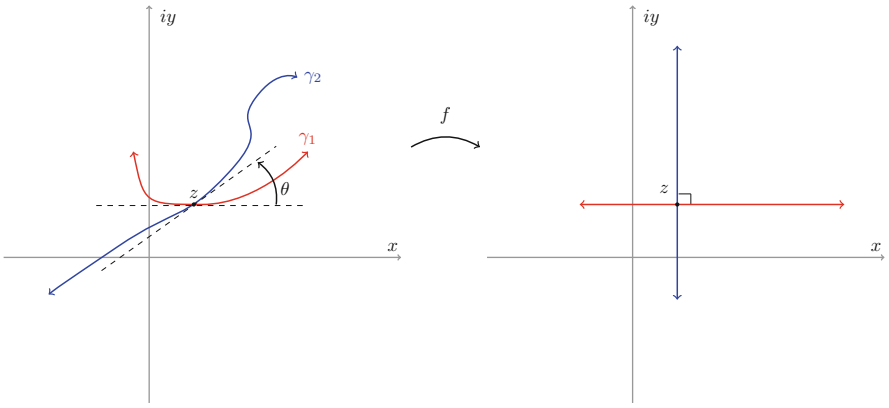



Figure 3.22: Not conformal

**Theorem 126.** *(Conformality Theorem) A holomorphic function $w = f(z)$ whose derivative at $z_0$ is nonzero will be conformal at $z_0$.*

*Proof.* Let the angle between $\gamma_1$ and $\gamma_2$ be $\theta = \arg \gamma_1'(z_0) - \arg \gamma_2'(z_0)$. The curves $\gamma_1$ and $\gamma_2$ are mapped by $f$ to paths $f \circ \gamma_1$ and $f \circ \gamma_2$; they meet at $f(z_0)$ at an angle

$$\phi = \arg(f \circ \gamma_1)'(z_0) - \arg(f \circ \gamma_2)'(z_0).$$

By the chain rule

$$(f \circ \gamma_1)'(z_0) = f'(\gamma_1(z_0))\,\gamma_1'(z_0) \quad \text{and} \quad (f \circ \gamma_2)'(z_0) = f'(\gamma_2(z_0))\,\gamma_2'(z_0).$$

Since both $\gamma_1$ and $\gamma_2$ are smooth, both $\gamma_1'$ and $\gamma_2'$ are nonzero. Furthermore, by the hypothesis $f'(z_0) \neq 0$ and $\gamma_1(z_0) = \gamma_2(z_0)$, the assertion of the theorem follows from

$$\frac{\gamma_1'(z_0)}{\gamma_2'(z_0)} = \frac{(f \circ \gamma_1)'(z_0)}{(f \circ \gamma_2)'(z_0)}$$

and the fact that $\arg\left(\dfrac{z}{w}\right) = \arg(z) - \arg(w)$. Thus $\phi = \theta$ in $[0, \pi]$. □

**Remark 70.** In the above proof, we could have used the scalar product given in (3.1) to prove the conformality theorem. Indeed if $\alpha = f'(z_0) \neq 0$, then it can be shown that

$$\langle \alpha z, \alpha w \rangle = |\alpha|^2 \langle z, w \rangle,$$

and thus the cosine of the angle between $\alpha z$ and $\alpha w$ is equal to the cosine of the angle between $z$ and $w$. The details are left to the reader.

**Example 101.**

a) The entire function $f(z) = e^z$ is conformal at every point in the complex plane since $f'(z) = e^z \neq 0$ for all $z \in \mathbb{C}$.

b) The entire function $g(z) = z^2$ is conformal when $z \neq 0$ since $g'(z) = 2z \neq 0$ when $z \neq 0$. However it is not conformal at $z = 0$; if we consider $f(z) = z^2 = w$ it takes the rays $\arg z = \theta$ and $\arg z = \mu$ meeting at the angle $\theta - \mu$ to rays $\arg w = 2\theta$ and $\arg w = 2\mu$ meeting at angle $2(\theta - \mu)$ as shown in Figure 3.23.

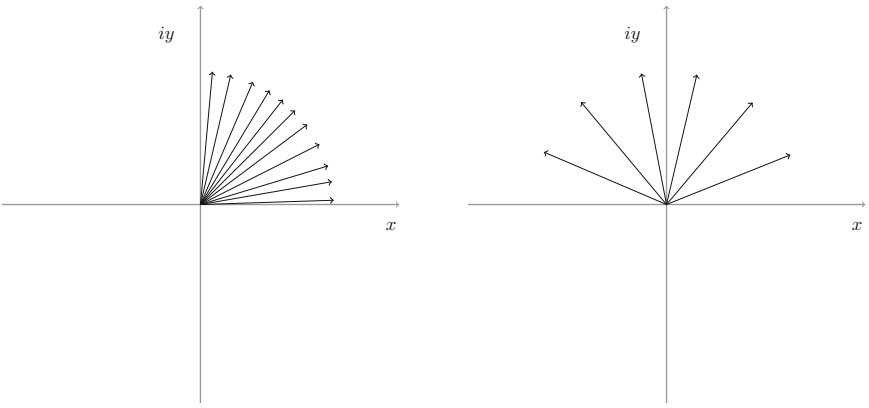



Figure 3.23: Non-conformality of $z \to z^2$ at 0

**Remark 71** (Angle magnification at a critical point)**.** If a complex-valued function $f$ is holomorphic at $z_0$ and if $f'(z_0) = 0$ then we say $z_0$ is a *critical point* for $f$. Although it does not follow from the above conformality theorem, it is true that holomorphic functions are not conformal at critical points due to some angle magnification at a critical point (see [54], p. 355). For example, the function $f(z) = \sin z$ is entire (i.e., holomorphic on the entire complex plane) and its derivative $f'(z) = \cos z = 0$ if and only if $z = (2n+1)\frac{\pi}{2}$ for $n = 0, \pm 1, \pm 2, \ldots$, so each of these point is a critical point of $f$. Thus $w = \sin z$ is a conformal mapping for all $z \neq (2n+1)\frac{\pi}{2}$, $n = 0, \pm 1, \pm 2, \ldots$.

## Conformal Equivalence: Riemann Mapping Theorem

Two domains $G$ and $G^*$ are said to be *conformally equivalent* if there is a one-to-one and onto (i.e., bijective) conformal map

$$f : G \to G^*.$$

If such function $f$ exists, then its inverse function will be holomorphic as well (Figure 3.24).

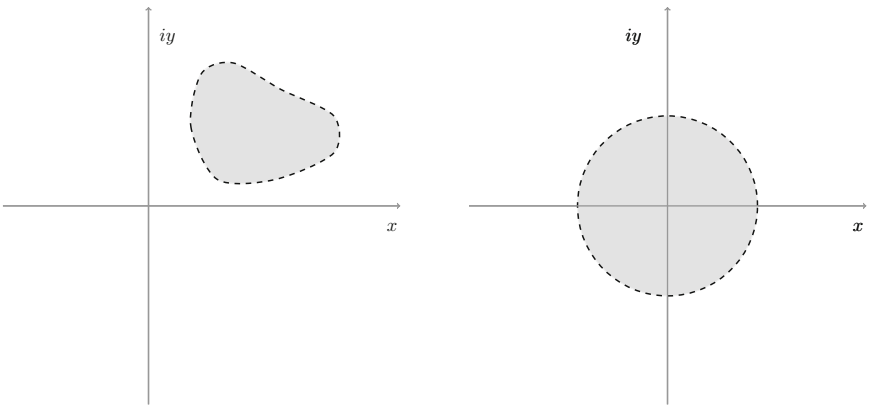



Figure 3.24: Conformally equivalent domains

In some cases when a bijection exists it can be given by explicit formulas. For example, in the following we will see that the upper half-plane can be mapped by a holomorphic bijection to the disc, and this is given by a linear fractional transformation. If two domains are conformally equivalent, then it is not possible to distinguish between them. The main theorem regarding conformability is the following Riemann Mapping Theorem.

**Definition 106.** A domain is called *simply connected* if every closed path $\gamma$ in $G$ is homotopic to a null path in $G$. Intuitively, this means if every closed path $\gamma$ in the domain can be continuously shrunk to a point.

**Theorem 127.** *(The Riemann Mapping Theorem) Two simply connected domains, neither of which is equivalent to* $\mathbb{C}$*, are conformally equivalent.*

This theorem, even though it is a corner stone result, is an existence result; it does not produce for us the desired function $f$. In practice, to find a conformal mapping from a domain $G$ onto a simpler domain $G^*$, we first find the conformal map from $G$ onto the open unit disc $|w| < 1$. We then apply the theorem once again to find a mapping $g$ from $G^*$ onto the open disc $|w| < 1$; since the theorem guarantees that $g$ is one-to-one, it has a inverse $g^{-1}$ that maps the open disc onto $G^*$. The desired mapping from $G$ onto $G^*$ is then given by $w = (g^{-1} \circ f)(z)$, i.e., to conformally relate the two domains, we are composing various maps to the disc and inverse maps to the disc. Note that it is not obvious that it is possible to construct a conformal mapping from a region with complicated boundary onto a nice region like the disc or vice versa. The Riemann Mapping Theorem is very striking in this context; it simply states that if $G$ is a simply connected region with $G \neq \mathbb{C}$, then there exists a one-to-one conformal mapping $f$ from $G$ onto the unit disc $\mathbb{D}$ with $f^{-1} : \mathbb{D} \to G$ also conformal. For the proof of the Riemann Mapping Theorem we refer the reader to [5], pp. 221–227.

## The Disc and Upper Half-Plane

The unit disc denoted by $\mathbb{D}$ is defined as $\mathbb{D} = \{z \in \mathbb{C} : \ |z| < 1\}$. The upper half-plane, which we denote by $\mathbb{H}$, consists of those complex numbers with positive imaginary parts; that is,

$$\mathbb{H} = \{z \in \mathbb{C} : \ \mathrm{Im}(z) > 0\}.$$

**Theorem 128.** *The map $F : \mathbb{H} \to \mathbb{D}$ defined by $F(z) = \dfrac{i-z}{i+z}$ is a conformal map with inverse $G : \mathbb{D} \to \mathbb{H}$ defined as $G(z) = i\dfrac{1-w}{1+w}$.*

*Proof.* Observe that $F$ maps $\mathbb{H}$ to $\mathbb{D}$ because $|i-z| < |i+z|$ and thus $|F(z)| < 1$. It is also easy to see that $G$ maps into the upper half-plane. Take $w = u + iv$ in $\mathbb{D}$ and observe that we have

$$\mathrm{Im}(G(w)) = \mathrm{Re}\left(\frac{1-w}{1+w}\right) = \frac{1-u^2-v^2}{(1+u)^2+v^2} > 0.$$

Lastly,

$$G(F(z)) = i\frac{1-\dfrac{i-z}{i+z}}{1+\dfrac{i-z}{i+z}} = z,$$

and similarly $F(G(w)) = w$. □

**Remark 72.** The above theorem at first seems surprising in the sense that $\mathbb{H}$ is an unbounded set which is conformally equivalent to the unit disc. This is one reason why complex analysts work on studying function theory on the disc; as seen above, knowledge about the disc can be carried out to knowledge about any simply connected region. However in the theory of several complex variables, the situation is different. There are many simply connected domains in $\mathbb{C}^n$ that are not conformally equivalent.

## Harmonic Functions and Conformal Maps

Laplace's equation

$$u_{xx} + u_{yy} = 0$$

is used often in electrostatics, heat conduction, and fluid flow. Harmonic functions which we introduce briefly below are used in the mathematical modeling of such phenomena. A classical problem involving Laplace's equation is called *the Dirichlet Problem*. Suppose $G$ is a domain in $\mathbb{R}^2$ and that $h$ is a function defined on the boundary of $G$. The problem of finding a function $u(x, y)$ which satisfies the Laplace's equation in $G$ and which equals $h$ on $\partial G$ (the boundary of $G$) is called a Dirichlet problem. Such problems arise often in the two-dimensional

modeling of heat flow, fluid flow, and electrostatics. Solving a Dirichlet problem by conformal mapping is often used. Roughly speaking, suppose that we have a region $G$ (a nonempty open connected set in $\mathbb{C}$), a one-to-one conformal mapping $g$ from $G$ onto the much simpler region $D(0,1)$

$$g : \overline{G} \to \overline{D(0;1)},$$

and $g$ maps the boundary $\partial G$ onto the unit circle. Since we have both $g$ and $g^{-1}$ are conformal (conformality of $g^{-1}$ is a consequence of the inverse function theorem), it turns out that they transfer harmonicity backward and forward. Hence the Dirichlet problem for $G$ is converted to a Dirichlet problem for $D(0,1)$. After solving the Dirichlet problem in $D(0,1)$, one transfers the solution to $G$.

**Definition 107.** A real-valued function $u(x,y)$ of two variables $x$ and $y$ that has continuous first and second order partial derivatives in a domain $D$ and satisfies Laplace's equation is said to be *harmonic* in $G$.

The following theorem shows that real and imaginary parts of a holomorphic functions are necessarily harmonic.

**Theorem 129.** *Let $f(z) = u(x,y) + iv(x,y)$ be holomorphic in an open set $G$. Then the functions $u(x,y)$ and $v(x,y)$ are harmonic in $G$.*

*Proof.* Since $f$ is holomorphic it is infinitely differentiable. Thus $u(x,y)$ and $v(x,y)$ have partial derivatives of all orders and so the partials are continuous. In particular $u$ and $v$ have continuous second order partials. Since $f$ is holomorphic, the Cauchy–Riemann equations are satisfied at every point $z$. By differentiating the both sides of $u_x = v_y$ and $u_y = -v_x$ with respect to $x$ and $y$ we find

$$u_{xx} = v_{yx} \quad \text{and} \quad u_{yy} = -v_{xy}.$$

Since we assumed continuity, the mixed partials $v_{xy}$ and $v_{yx}$ are equal. Hence adding the two equations above we obtain

$$u_{xx} + u_{yy} = 0.$$

This shows $u(x,y)$ is harmonic. Similarly by taking $u_x = v_y$, differentiating both sides with respect to $y$ and differentiating both sides of $u_y = -v_x$ with respect to $x$ gives

$$u_{xy} = v_{yy} \quad \text{and} \quad u_{yx} = -v_{xx}.$$

Subtracting the last two equations will yield $v_{xx} + v_{yy} = 0$. □

Now the question is if you are given a harmonic function $u(x,y)$ in $G$, is it possible to find another harmonic function $v(x,y)$ so that $u$ and $v$ satisfy the Cauchy–Riemann equations throughout the domain $G$? In this case we call $v(x,y)$ a *harmonic conjugate* of $u(x,y)$. By combining the functions $u(x,y)$ and $v(x,y)$ as $u(x,y) + iv(x,y)$ we obtain a holomorphic function in $G$.

**Example 102.** The functions $u(x,y) = x^2 - y^2$ and $v(x,y) = 2xy$ are harmonic conjugates of each other, since they are the real and imaginary parts of $f(z) = z^2$.

**Example 103.** Given one of the two conjugate harmonic functions, the Cauchy–Riemann equations can be used to find the other. Consider $u(x,y) = y^3 - 3x^2y$. We observe that $u(x,y)$ is harmonic because $u_x = -6xy$, $u_{xx} = -6y$, $u_y = 3y^2 - 3x^2$, $u_{yy} = 6y$ and hence $u_{xx} + u_{yy} = 0$. In order to find the harmonic conjugate $v(x,y)$, note that

$$u_x = -6xy = v_y.$$

Partial integration of $-6xy = v_y$ with respect to $y$ will give $v(x,y) = -3xy^2 + \phi(x)$ and since $v_x = -u_y$, we also have

$$-3y^2 + \phi'(x) = -(3y^2 - 3x^2).$$

So we obtain $\phi'(x) = 3x^2$ and thus $\phi(x) = x^3 + C$, where $C$ is a real constant. Therefore, the harmonic conjugate of $u(x,y)$ is

$$v(x,y) = -3xy^2 + x^3 + C.$$

The corresponding function $f = u + iv$ is $f(z) = (y^3 - 3x^2y) + i(x^3 - 3xy^2) + iC$, and one can verify that $f(z) = i(z^3 + C)$.

## Linear Fractional Transformations

A map of the form $f(z) = az + b$ where $a$ and $b$ are complex numbers with $a \neq 0$ is called a *linear transformation.* It is easy to see that every non-constant complex linear transformation can be described as a composition of three basic types of motion. Namely, a translation, a rotation and a magnification. To see this all we need to do is to write $a = \rho\, e^{i\phi}$. Then

$$w = f(z) = \rho\, e^{i\phi} z + b$$

and $e^{i\phi}\, z$ is a rotation, $\rho\, e^{i\phi} z$ is a magnification, and $\rho e^{i\phi} z + b$ is a translation. For example the linear transformation

$$w = f(z) = 2iz + 1 + i = 2e^{i\frac{\pi}{2}}\, z + (1 + i)$$

is a rotation by $\frac{\pi}{2}$, magnification by 2 and a translation by $1 + i$.
An *inversion* is a transformation defined by

$$w = f(z) = \frac{1}{z}.$$

It is a one-to-one mapping of the extended complex plane onto itself and turns the unit circle inside out. Observe that

$$w = f(z) = \frac{1}{r\, e^{i\phi}} = \frac{1}{r}\, e^{-i\phi}.$$

Thus, if we set $r$ fixed and $\phi$ varies from 0 to $2\pi$ then the circle of radius $r$ centered at the origin is mapped onto a circle of radius $\frac{1}{r}$ centered at the origin and traced backward. On the other hand, if we set $\phi$ fixed and let the real number $r$ vary from $-\infty$ to $\infty$ then the point at $\infty$ goes to the origin and vice versa. A line passing through the origin making an angle $\phi$ with the positive $x$-axis in the $z$-plane is mapped to a line passing through the origin making an angle $-\phi$ with the positive $x$-axis in the $w$-plane.

However, it is wrong to conclude that inversion maps lines to lines and circles to circles. But it does the next best thing: the image of a line is either a line or a circle. By treating $\infty$ as a point we could regard a line as a *generalized circle* with radius $\infty$ and can say an inversion maps generalized circles to generalized circles (one can prove this statement using stereographic projection). At any rate, it is not difficult to check whether the image will be a line or a circle.

a) If the original line or circle passes through the origin, its image will contain the point at infinity, so it must be a line.

b) If the original line or circle misses the origin, its image will be bounded, hence a circle.

**Definition 108.** A linear fractional transformation (or Möbius transformation, or bilinear transformation) $T : \tilde{\mathbb{C}} \to \tilde{\mathbb{C}}$ is a function of the form

$$T(z) = \frac{az + b}{cz + d}$$

with the restriction $ad \neq bc$ and $a, b, c, d \in \mathbb{C}$.

The following are some properties of linear fractional transformations that can be deduced from the definition.

a) When $c = 0$, the Möbius transformation

$$T(z) = \frac{az + b}{d} = \frac{a}{d}z + \frac{b}{d}$$

is a linear transformation. If $c \neq 0$ then after some algebra one can show

$$T(z) = \frac{a}{c} + \frac{bc - ad}{c} \frac{1}{cz + d}.$$

Hence $T$ can be expressed as a composition of a finite sequence of translations, magnifications, rotation, and inversions.

b) We may arrange these numbers $a, b, c, d \in \mathbb{C}$ appearing in the above definition as a matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ for which $\det A \neq 0$. Thus, $A$ is invertible, in fact $A^{-1} = \dfrac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.

Moreover, if $\lambda \neq 0$, $\lambda \in \mathbb{C}$, then the matrix $B = \begin{bmatrix} \lambda a & \lambda b \\ \lambda c & \lambda d \end{bmatrix}$ gives rise to the same map.

c) $T'(z) \neq 0$ for all $z \neq -\frac{d}{c}$, $T : \tilde{\mathbb{C}} \to \tilde{\mathbb{C}}$ is a one-to-one, onto map, and its inverse can be found either by setting $w = T(z) = \frac{az+b}{cz+d}$ and solving for $z$, or simply using the matrix $\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$ to write $z = T^{-1}(z) = \frac{dw-b}{-cw+a}$.

d) From $T(z) = \frac{az+b}{cz+d}$ , we conclude

$$T\left(\frac{-b}{a}\right) = 0, \quad T\left(\frac{-d}{c}\right) = \infty, \quad \lim_{z\to\infty} T(z) = \frac{a}{c}, \quad \text{and} \quad T(0) = \frac{b}{d}.$$

The above set of equalities can be used to find a particular linear fractional transformation. For example, if we are asked to find the linear fractional transformation which maps the points $0, 1, \infty$ onto $1, \infty, 0$, we simply observe $b = d$, $\quad c = -d$ because $T(0) = \frac{b}{d} = 1$ and $T(1) = \infty$, so $\frac{-d}{c} = 1$. Moreover $T(\infty) = \frac{a}{c} = 0$ implies $a = 0$. Putting all of this information together will yield $T(z) = \frac{1}{-z+1}$.

e) Recall that a point $z_0$ is a fixed point of $T$ if $T(z_0) = z_0$. Given a linear fractional transformation $T$ , if $\infty$ is a fixed point of $T$, then there exists complex numbers $a, b$ such that $T(z) = az + b$. This follows from the fact that $\lim_{z\to\infty} T(z) = \frac{a}{c}$, by hypothesis $c = 0$, and

$$T(z) = \frac{az+b}{d} = \frac{a}{d}z + \frac{b}{d}.$$

Perhaps it is not so straightforward to see that given a linear fractional transformation, if $T$ has three fixed points then $T$ is the identity (see Exercise 6 below).

## Conformality of Möbius Transformation

If we think of the Möbius transformation as a map taking $z$ to $T(z) = \frac{az+b}{cz+d}$ with $ad \neq bc$ , then $T'(z) = \frac{ad-bc}{(cz+d)^2} \neq 0$ and therefore $T$ is conformal in

$\mathbb{C} \setminus \left\{ \dfrac{-d}{c} \right\}$ for $c \neq 0$. However, we defined $T : \tilde{\mathbb{C}} \to \tilde{\mathbb{C}}$ where $\tilde{\mathbb{C}} = \mathbb{C} \cup \infty$. Thus we need to extend the definition of conformality to such mappings. We say $T$ is conformal at $\infty$ if $T^*(z) = T(1/z)$ is conformal at $z = 0$. If $T$ maps $\varsigma \in \mathbb{C}$ to $\infty$, then we set $S(z) = \dfrac{1}{T(z)}$ and we say $T$ is conformal at $\infty$ if $S$ is conformal at $\varsigma$. Another way to think about $T$ being conformal at $\infty$ is to define $T^*$ as $T^*(z) = f\left(\dfrac{1}{z}\right)$ and check if $T^*$ is conformal at 0. With a little more work, one can show that the Möbius transformation has nonzero derivative at every point of $\tilde{\mathbb{C}}$. Consider cases separately for $c = 0$ and $c \neq 0$. For example, for the case $c \neq 0$, since we know $T(-d/c) = \infty$, by taking $w = T(z)$ and letting $\tau = 1/w$, we get $\tau = \dfrac{cz + d}{az + b}$ and this function has a nonzero derivative at $z = 0$.

**Theorem 130.** *Given any three distinct points $z_1$, $z_2$, $z_3$ on the Riemann sphere and three distinct points $w_1$, $w_2$, $w_3$ there exists a unique fractional linear transformation $T$ such that*

$$T(z_i) = w_i \quad \textit{for} \quad i = 1, 2, 3.$$

*Proof.* Suppose there are two linear fractional transformations $T$ and $S$ such that $T(z_i) = w_i$ and $S(z_i) = w_i$ for $i = 1, 2, 3$. Then the mapping

$$(T \circ S^{-1})(w_i) = T(z_i) = w_i \quad \text{for} \quad i = 1, 2, 3,$$

i.e., $T \circ S^{-1}$ has three fixed points, then $T \circ S^{-1} = I$ or $T = S$. □

## Triplet Representation of Möbius Transformation

If we want to give an expression to a function which maps $z_1$ to 0 and $z_2$ to $\infty$ then the function $T(z) = \dfrac{z - z_1}{z - z_2}$ will do the job. Suppose we also want a third point $z_3$ to be mapped to 1 as well, then the map

$$T(z) = \frac{z - z_1}{z - z_2} \frac{z_3 - z_2}{z_3 - z_1}$$

is the right one. Actually, we will know more when we express $T(z)$ in the next theorem.

**Theorem 131.** *The function*

$$z \to \frac{z - z_1}{z - z_2} \frac{z_3 - z_2}{z_3 - z_1}$$

*is the unique function such that $T(z_1) = 0, T(z_2) = \infty$, and $T(z_3) = 1$.*

**Definition 109.** If $w = T(z)$ is a function such that $T(z_i) = w_i$ for $i = 1, 2, 3$, then $w$ and $z$ are related by the *cross-ratio formula* given as:

$$\frac{w - w_1}{w - w_2}\frac{w_3 - w_2}{w_3 - w_1} = \frac{z - z_1}{z - z_2}\frac{z_3 - z_2}{z_3 - z_1}.$$

**Remark 73.** Notice that in case $T(z_1) = 0$, $T(z_2) = 1$ and $T(z_3) = \infty$, then the cross-ratio formula will be

$$\frac{z - z_1}{z - z_2}\frac{z_3 - z_2}{z_3 - z_1}.$$

Furthermore, the cross-ratio formula can be modified so that one can still use it when $\infty$ is one of the prescribed points in the extended $z$ or $w$ plane. For example if we suppose $z_1 = \infty$, then

$$\lim_{z_1 \to \infty} \frac{z - z_1}{z - z_2}\frac{z_3 - z_2}{z_3 - z_1} = \frac{z_3 - z_2}{z - z_2}.$$

Thus the cross-ratio formula takes the form

$$\frac{w - w_1}{w - w_2}\frac{w_3 - w_2}{w_3 - w_1} = \frac{z_3 - z_2}{z - z_2}.$$

**Example 104.** Find a linear fractional transformation $T$ which maps $0, 1, \infty$ onto $-1, 0, 1$. In this case the triplet representation of $T$ is:

$$\frac{w - w_1}{w - w_2}\frac{w_3 - w_2}{w_3 - w_1} = \frac{z - z_1}{z - z_2}$$

and therefore

$$\frac{w + 1}{w} \cdot \frac{1 - 0}{2} = \frac{z - 0}{z - 1}$$

and $w = \dfrac{z - 1}{z + 1}$.

## Some Standard Examples

In the following we give some examples to illustrate various isomorphisms; details are left to the reader.

a) The quarter disc and the half disc (Figure 3.25).

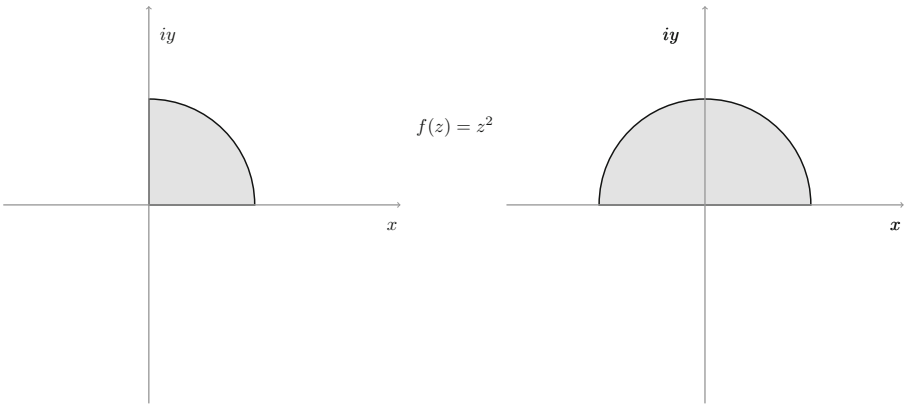



Figure 3.25: $z \mapsto z^2$

b) First quadrant and upper half-plane (Figure 3.26).

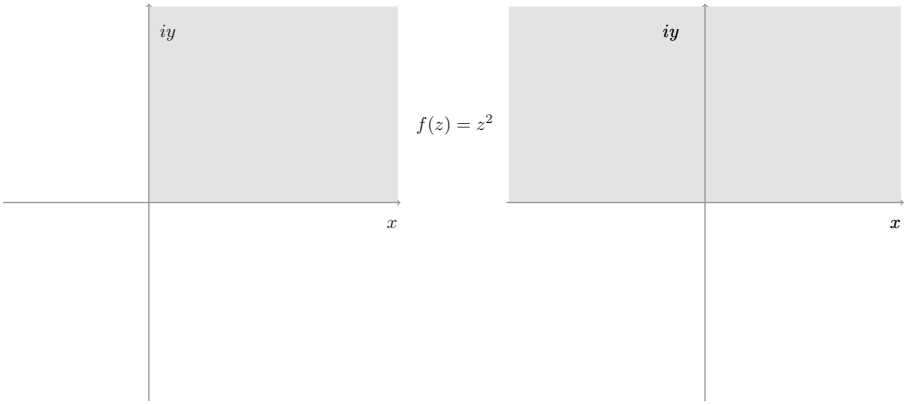



Figure 3.26: $z \mapsto z^2$

c) Upper half disc and first quadrant (Figure 3.27).

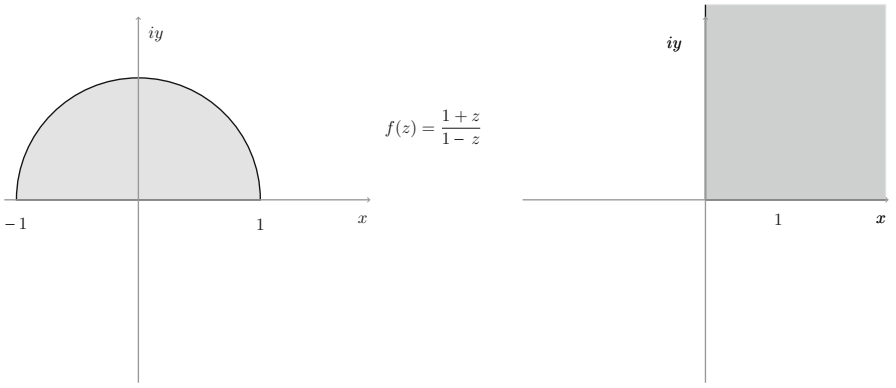



Figure 3.27: $z \mapsto \frac{1+z}{1-z}$

d) Upper half-strip with upper half-plane (Figure 3.28).

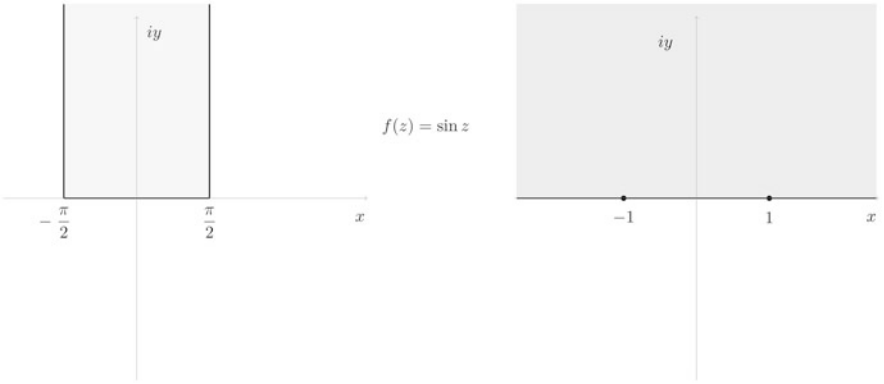



Figure 3.28: $z \mapsto \sin z$

e) Upper half disc with a half-strip (Figure 3.29).

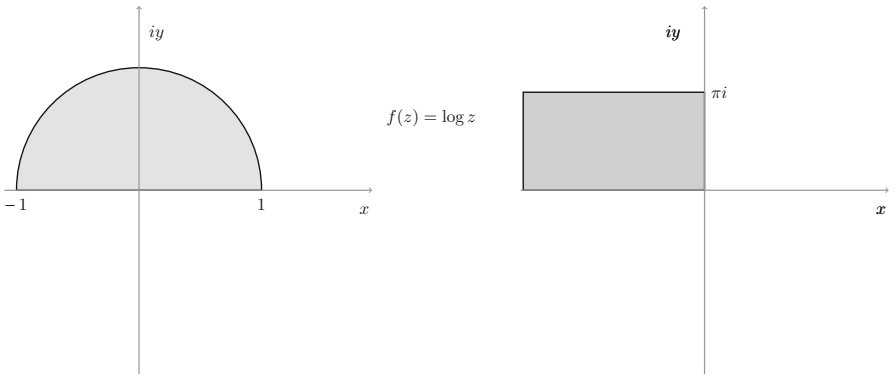



Figure 3.29: $z \mapsto \log z$

f) The quarter disc and upper half-plane (Figure 3.30). Note that the above

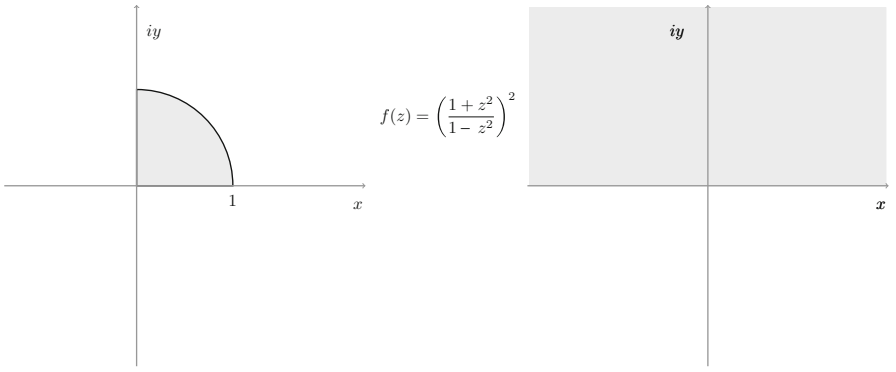



Figure 3.30: $z \mapsto \left(\frac{1+z^2}{1-z^2}\right)^2$

example $z \mapsto \left(\frac{1+z^2}{1-z^2}\right)^2$ is a composition of three maps. First consider the map $z \mapsto z^2$ between the quarter disc and half disc in example (a), then compose by the map $f(z) = \dfrac{1+z}{1-z}$ in example (c) followed by the map $f(z) = z^2$ in example (b).

## Exercises

1. Find a conformal map taking $G = \{z : \ 0 < \arg z < \frac{\pi}{2}, \quad 0 < |z| < 1\}$ to $D = \{z : \ |z| < 1\}$.

2. Find a conformal map from $\{z \in \mathbb{C} : |z| < 1, \text{Re}z > 0\}$ onto $D\{z \in \mathbb{C} : |z| < 1\}$.

3. Let $f(z) = \dfrac{z-1}{z+1}$. What is the image under $f$ of

a) The real line?

b) The circle with center 0 radius 2?

c) The imaginary axis?

4. Find a linear transformation mapping the circle $|z| = 1$ onto the circle $|w - 5| = 3$ and taking the point $z = i$ to $w = 2$.

5. Suppose the linear transformation is given by $w = (1 + i)z + (1 - i)$. What is the image of the rectangular region defined by $O = (0, 0), A = 1$ and $B = 2i$?

6. A *fixed point* of a function $f(z)$ is a point $z_0$ satisfying $f(z_0) = z_0$. Show that a Möbius transformation $f(z)$ can have at most two fixed points in the complex plane unless $f(z) \equiv z$.

7. Let $T$ be a linear fractional transformation. Prove that if $T$ has three fixed points then $T$ is an identity map.
Hint: Consider the cases where $\infty$ is one of the fixed points and $\infty$ is not a fixed point for $T$.
8. Find the Möbius transformation that maps $0, 1, \infty$ to the following respective points:

a) $0, i, \infty$

b) $0, 1, 2$

c) $-i, \infty, 1$

d) $-1, \infty, 1$

9. Discuss the image of the circle $|z - 2| = 1$ and its interior under the following transformations:

a) $w = z - 2i$

b) $w = 3iz$

c) $w = \dfrac{z-2}{z-1}$

d) $w = \dfrac{z-4}{z-3}$

e) $w = \dfrac{1}{z}$

10. Prove that any linear fractional transformation $T$ with $c \neq 0$ can be written as:

$$T = T_4 \circ T_3 \circ T_2 \circ T_1,$$

where

$$T_1(z) = z + \frac{d}{c}, \quad T_2(z) = \frac{1}{z}, \quad T_3(z) = \left[\frac{(bc - ad)}{c^2}\right] z \quad \text{and} \quad T_4(z) = z + \frac{a}{c}.$$

Interpret $T$ geometrically.

11. Show that the map $w = \overline{z}$ is not a conformal mapping at the point $z_0 = 1+i$.

12. Use the scalar product given in (3.1) to prove Theorem 126.

13. Show that if $F$ and $G$ are linear fractional maps, then so is $F \circ G$.

## 3.6 Integration in the Complex Plane

### Integration Along Paths

It is well known that continuous real-valued functions on compact intervals in $\mathbb{R}$ are integrable. However, continuous functions are not so adequate for the integration of complex-valued functions. We say that a real or complex-valued function $h$ is *piecewise continuous* on a compact interval $[\alpha, \beta] \subset \mathbb{R}$ if there exists points

$$\alpha = t_0 < t_1 < \cdots < t_n = \beta$$

and continuous functions $h_k$ on $[t_k, t_{k+1}]$ such that $h_k(t) = h(t)$ for $t \in (t_k, t_{k+1})$ for $k = 0, \ldots, n-1$. Basically this means that $h$ is continuous except possibly for a finite number of jump discontinuities. A real-valued piecewise continuous function $h$ is integrable with

$$\int_\alpha^\beta h(t)\, dt = \sum_{k=0}^{n-1} \int_{t_k}^{t_{k+1}} h_k(t)\, dt.$$

On the other hand, if $g$ is a complex-valued function defined on $[\alpha, \beta] \subset \mathbb{R}$ where $g = \operatorname{Re} g + i \operatorname{Im} g$, we say $g$ is integrable if and only if $\operatorname{Re} g$ and $\operatorname{Im} g$ are both integrable and we define

$$\int_\alpha^\beta g(t)dt := \int_\alpha^\beta \operatorname{Re} g(t)dt + i \int_\alpha^\beta \operatorname{Im} g(t)dt.$$

For example,

$$\int_0^{2\pi} e^{it} dt = \int_0^{2\pi} \cos t\, dt + i \int_0^{2\pi} \sin t\, dt = \sin t|_0^{2\pi} + i[-\cos t]|_0^{2\pi} = 0.$$

**Definition 110.** Let $[\alpha, \beta]\,(-\infty < \alpha \leq \beta < \infty)$ be a closed and bounded interval in $\mathbb{R}$. A *curve* $\gamma$ with parameter interval $[\alpha, \beta]$ is a continuous function

$$\gamma : [\alpha, \beta] \to \mathbb{C}$$

with initial point $\gamma(\alpha)$ and terminal point $\gamma(\beta)$. The curve $\gamma$ is *closed* if $\gamma(\alpha) = \gamma(\beta)$. It is *simple* if it does not intersect itself ( if $\alpha \leq t_1 < t_2 \leq \beta$ implies that $\gamma(t_1) \neq \gamma(t_2)$ unless $t_1 = t_2$). We denote the image by $\gamma^* = \gamma([\alpha, \beta])$. Notice that as a continuous image of a compact interval, $\gamma^*$ is a compact subset of $\mathbb{C}$.

The curve $\gamma$ has a certain orientation determined as $t$ increases from $\alpha$ to $\beta$. Given $\gamma$, there exists a curve $-\gamma$ with the same image set but the opposite orientation:

$$-\gamma(t) := \gamma(\alpha + \beta - t) \quad \text{where} \quad t \in [\alpha, \beta].$$

We are now ready to define the integral of a function along a path: let $\gamma$ be a path with parameter interval $[\alpha, \beta]$. Suppose $\alpha = t_0 < t_1 < \cdots < t_n = \beta$ such that $\gamma$ restricted to each interval $[t_k, t_{k+1}]$ is a continuously differentiable function. Let $f : \gamma^* \to \mathbb{C}$ be continuous. Then the *integral of $f$ along $\gamma$* or round $\gamma$ if $\gamma$ is closed is defined as:

$$\int_\gamma f(z)\,dz := \int_\alpha^\beta f(\gamma(t))\gamma'(t)\,dt.$$

Note that since

$$(f \circ \gamma)\gamma' : t \mapsto f(\gamma(t))\gamma'(t)$$

is piecewise continuous and hence integrable, and the above integral on the right makes sense. Given $f(z) = u(z) + iv(z)$, if we want to know the relationship between the line integrals of $u(x(t), y(t))$ and $v(x(t), y(t))$ along $\gamma$ and the complex integral $\int_\gamma f(z)dz$ we simply write:

$$\begin{aligned}\int_\gamma f(z)dz &= \int_\alpha^\beta (u + iv)(x' + iy')dt \\ &= \int_\alpha^\beta (ux' - vy')dt + i\int_\alpha^\beta (uy' + vx')dt \\ &= \int_\gamma udx - \int_\gamma vdy + i\int_\gamma vdx + i\int_\gamma udy.\end{aligned}$$

## Basic Properties of Integrals Along Paths

- Join of paths: Suppose $\gamma$ is a path with parameter interval $[\alpha, \beta]$ and is the join of paths $\gamma_1, \gamma_2, \ldots, \gamma_n$, and let $f : \gamma^* \to \mathbb{C}$ be continuous. Then

$$\int_\gamma f(z)\,dz = \sum_{k=1}^n \int_{\gamma_k} f(z)\,dz.$$

- Reversal: Suppose $\gamma$ is a path with parameter interval $[\alpha, \beta]$ and let $f : \gamma^* \to \mathbb{C}$ be continuous. Then

$$\int_{-\gamma} f(z)dz = -\int_{\gamma} f(z)\, dz.$$

- Reparametrization: Let $\tilde{\gamma}$ be another path with parameter interval $[\tilde{\alpha}, \tilde{\beta}]$ and $g$ be a function which maps $[\tilde{\alpha}, \tilde{\beta}]$ onto $[\alpha, \beta]$; moreover assume $g$ has a positive continuous derivative. Then

$$\int_{\gamma} f(z)\, dz = \int_{\tilde{\gamma}} f(z)\, dz.$$

Proofs of the above properties can be found in any standard book on complex analysis, e.g., [37], pp. 94–111. What we will be interested is the path integrals of holomorphic functions around simple closed curves oriented counterclockwise. The following easy examples motivate our main theorem.

**Example 105.** 1. Suppose we have $f(z) = z$ and our simple closed curve $\gamma$ is the unit circle $\gamma(0, 1) = e^{it} = |z| = 1$ oriented counterclockwise. Then,

$$\int_{\gamma} f(z)\, dz = \int_{\gamma} z\, dz = \int_0^{2\pi} e^{it}\, ie^{it} dt = 0.$$

2. Suppose we have $f(z) = \dfrac{1}{z}$ and our simple closed curve $\gamma$ is the unit circle $\gamma(0, 1) = e^{it} = |z| = 1$ oriented counterclockwise. Since $|z|^2 = z\,\overline{z} = 1$, our function $f(z) = \dfrac{1}{z} = \overline{z}$; hence

$$\int_{\gamma} f(z)\, dz = \int_{\gamma} \overline{z}\, dz = \int_0^{2\pi} (\cos t - i \sin t) ie^{it} dt = 2\pi i.$$

We will notice soon the reason $\displaystyle\int_{\gamma} \frac{1}{z} = 2\pi i$ for the unit circle is that the function $f(z) = \dfrac{1}{z}$ is not holomorphic in the interior of the unit circle, namely at $z = 0$, otherwise the integral would be zero. The integral in the next example is often called the *fundamental integral* and it combines and generalizes the two integrals given in the above example.

**Example 106.** Let $a \in \mathbb{C}$ and $r > 0$, then we claim:

$$\int_{\gamma} (z - a)^n dz = \begin{cases} 0 & \text{if } n \neq -1 \\ 2\pi i & \text{if } n = -1 \end{cases}.$$

Here $\gamma = \gamma(a, r)$ denotes the circle centered at $a$ with radius $r$. Since for $t \in [0, 2\pi]$, $\gamma(a, r) = a + re^{it}$, using the very definition of the integral along a path we get

$$\begin{aligned}\int_\gamma (z-a)^n dz &= \int_0^{2\pi} (re^{it})^n rie^{it} dt \\ &= ir^{n+1} \int_0^{2\pi} e^{i(n+1)t} dt \\ &= ir^{n+1} \int_0^{2\pi} [\cos(n+1)t + i\sin(n+1)t] dt.\end{aligned}$$

For the case $n = -1$ the above integral $\int_\gamma (z-a)^n dz = i\int_0^{2\pi} \cos 0t dt = i2\pi$ and for the case $n \neq -1$

$$\int_\gamma (z-a)^n dz = ir^{n+1} \left( \left[ \frac{\sin(n+1)t}{n+1} \right]_0^{2\pi} - i \left[ \frac{\cos(n+1)t}{n+1} \right]_0^{2\pi} \right) = 0.$$

## The Complex Fundamental Theorem of Calculus

Recall that by the Fundamental Theorem of Calculus for real integrals, if $F(x)$ is an antiderivative of a function $f$, that is $F$ is a continuous function for which $F'(x) = f(x)$, then the integral of $f$ on the interval $[a, b]$ is the number

$$\int_a^b f(x)\, dx = F(x)|_a^b = F(b) - F(a).$$

Suppose $\gamma$ is a path with parameter interval $[\alpha, \beta]$ and that $G(z)$ is a function defined on an open set which also contains $\gamma^* = \gamma([\alpha, \beta])$. Suppose $G'(z)$ exists and is continuous at each point of $\gamma^*$. Then

$$\int_\gamma G'(z)\, dz = \begin{cases} G(\gamma(\beta)) - G(\gamma(\alpha)) & \text{in general} \\ 0 & \text{if } \gamma \text{ is closed.} \end{cases}$$

Clearly, this is an analog of the Fundamental Theorem of Calculus for real integrals. However, it is not as "fundamental" as in the real case. It is rather a stepping stone to Cauchy's Theorem and its consequences. For the proof of the Complex Fundamental Theorem of Calculus we refer the reader to [41], p. 124.

**Example.** Let $\gamma(t) = e^{it}$ for $0 \leq t \leq 2\pi$ and $f(z) = \operatorname{cosec}^2 z$; we know $f(z) = \dfrac{d}{dz}(-\cot z)$ in an open set containing $\gamma(t)$. Thus, $\int_\gamma \operatorname{cosec}^2 z\, dz = 0$ by the Complex Fundamental Theorem of Calculus.

If a complex integral cannot be evaluated explicitly, then one might like to get an estimate for it. For that purpose, the following lemma is very useful.

## Estimation Lemma

**Lemma 16.** *Suppose $f : \gamma^* \to \mathbb{C}$ is continuous, where $\gamma$ is a path with parameter interval $[\alpha, \beta]$ and $\gamma^* = \gamma([\alpha, \beta])$. Then*

$$\left| \int_\gamma f dz \right| \leq \int_\alpha^\beta |f(z)||dz|.$$

*In particular, if there exists a real number $M$ such that $|f(z)| \leq M$ for all $z \in \gamma^*$, then*

$$\left| \int_\gamma f(z) dz \right| \leq ML,$$

*where $L$ denotes the length of $\gamma$, which is, by definition,*

$$L = \int_\alpha^\beta |\gamma'(t)| dt.$$

Note that for circular arcs, line segments and therefore for contours, the definition for $L$ is exactly what we expect the "length" to be.

*Proof.* The idea of the proof follows from the fact that

$$\left| \int_\gamma f(z)\, dz \right| \leq \int_\alpha^\beta |f(\gamma(t)\, \gamma'(t)|\ dt.$$

If $|f(z)| \leq M$ for all $t \in \gamma^*$ then

$$\left| \int_\gamma f(z)\, dz \right| \leq M \int_\alpha^\beta |\gamma'(t)|\ dt. \qquad \square$$

## Exercises

1. Define parametrically a path $\gamma$ for which $\gamma^*$ is:

   a) The square with vertices $1 + i$, $-1 + i$, $-1 - i$, $1 - i$.

   b) The pair of circles $|z - 1| = 1$ and $|z + 1| = 1$, one traced clockwise and the other counterclockwise.

   c) The closed semicircle in the right half-plane with $i[-R, R]$ as diameter.

2. Describe the image $\gamma^*$ of $\gamma$ in the following:

   a) $\gamma(t) = 1 + ie^{it}, \quad t \in [0, \pi]$.

   b) $\gamma(t) = e^{it}, \quad t \in [-\pi, 2\pi]$.

c)

$$\gamma(t) = \begin{cases} tiz_0 + (1-t)z_0 & 0 \le t \le 1, \\ (t-1)(-z_0) + (2-t)iz_0 & 1 \le t \le 2, \\ (t-2)(-iz_0) + (3-t)(-z_0) & 2 \le t \le 3, \\ (t-3)z_0 + (4-t)(-iz_0) & 3 \le t \le 4. \end{cases}$$

3. Does $\operatorname{Re}\left(\int_\gamma f(z)\,dz\right) = \int_\gamma \operatorname{Re}\ f(z)\,dz$ ?
4. Evaluate the following integrals:

a) $\displaystyle\int_{|z|=1} \frac{dz}{z}$

b) $\displaystyle\int_{|z|=1} \frac{dz}{|z|}$

c) $\displaystyle\int_{|z|=1} \frac{|dz|}{z}$

d) $\displaystyle\int_{|z|=1} \frac{|dz|}{|z|}$

5. Evaluate the following integrals:

a) $\displaystyle\int_\gamma x\,dz$ where $\gamma$ is the circumference of the unit square.

b) $\displaystyle\int_\gamma e^z dz$ where $\gamma$ is the part of the unit circle joining 1 to $i$ in a counterclockwise direction.

6. Compute

$$\int_\gamma \left(z^2 - 3|z| + \operatorname{Im} z\right) dz,$$

where $\gamma$ is the quarter circle centered at the origin extending from 2 to $2i$.

7. Evaluate

$$\int_\gamma \frac{1}{z^2 - 2i}\,dz$$

where $\gamma$ is the right half of the circle $|z| = 6$ from $z = -6i$ to $z = 6i$.
8. Let $\gamma$ be the upper half of the unit circle described counterclockwise. Show that

$$\left|\int_\gamma \frac{e^z}{z}\right| \le \pi e.$$

9. Estimate the absolute value of $\int_\gamma \frac{dz}{2+z^2}$ where $\gamma$ is the upper half of the unit circle.

10. Let $\gamma$ be a unit circle. Prove that

$$\left| \int_\gamma \frac{\sin z}{z^2} dz \right| \leq 2\pi e.$$

## 3.7 Cauchy's Theorem

### Cauchy's Theorem: Basic Version

One form of Cauchy's theorem states that if $\gamma$ is a simple closed curve and if $f$ is holomorphic inside $\gamma$ and on $\gamma$ then $\int_\gamma f = 0$. This theorem is very central to the theory of holomorphic functions. If the function is not holomorphic on the whole region inside $\gamma$ then the integral may or may not be zero. For example, $f(z) = \frac{1}{z}$ on the unit circle $\gamma$ is holomorphic except at $z = 0$ and

$$\int_\gamma \frac{1}{z}\, dz = \int_0^{2\pi} \frac{1}{e^{it}} e^{it} i dt = 2\pi i,$$

while $f(z) = \frac{1}{z^2}$ is still holomorphic except at $z = 0$ and

$$\int_\gamma \frac{1}{z^2}\, dz = \int_0^{2\pi} \frac{1}{e^{2it}} e^{it} i dt = 0.$$

As stated above, Cauchy's theorem demonstrates that the path integral of a holomorphic function around a closed curve will be zero provided that the function is also holomorphic on the interior of the curve. To prove the preliminary version of Cauchy's theorem, we need to recall Green's Theorem which we state below.

**Theorem 132** (Green's Theorem)**.** *Let $\gamma$ be a simple closed curve oriented counterclockwise in $\mathbb{C}$ and $\Omega$ be its interior. If $P(x,y)$ and $Q(x,y)$ are two real-valued differentiable functions, then*

$$\int_\gamma P(x,y)dx + Q(x,y)dy = \iint_\Omega \left( \frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx\, dy.$$

For the following theorem recall that by $H(G)$ we mean the set of holomorphic functions in $G$.

**Theorem 133** (Preliminary version of Cauchy's Theorem)**.** *Let $G$ be a simply connected region and $f \in H(G)$ with derivative $f'(z)$ continuous on and inside a simple closed curve $\gamma$ in $G$. Then*

$$\int_\gamma f(z)dz = 0$$

*on every simple closed curve $\gamma$ in $G$.*

*Proof.* Let $f(z) = u(z) + iv(z)$, with $u(z), v(z)$ real-valued functions; we know that Cauchy–Riemann equations hold: $u_x = v_y \quad \text{and} \quad u_y = -v_x$. Then

$$\begin{aligned}\int_\gamma f(z)dz &= \int_\gamma (u+iv)(dx+idy) \\ &= \int_\gamma (udx - vdy) + i\int_\gamma (vdx + udy). \\ &= \iint_\gamma (-v_x - u_y)dxdy + i\iint_\gamma (u_x - v_y)dxdy.\end{aligned}$$

Both terms are zero by the Cauchy–Riemann Equations. □

**Definition 111.** We define a **contour** to be a simple closed path whose image is made up of a finite number of line segments and circular arcs. The positive direction on a contour $\gamma$ is the direction corresponding to increasing values of the parameter $t$. The positive direction roughly corresponds to the direction that a person must walk on $\gamma$ in order to keep the interior of $\gamma$ to the left. For example, the circle $z(t) = e^{it}$ for $0 \le t \le 2\pi$ has a positive orientation. The negative direction on a contour $\gamma$ is the direction opposite to the positive direction. If $\gamma$ has an orientation, the opposite curve, that is a curve with opposite orientation, is denoted by $-\gamma$. The purpose of introducing contours is to bypass the geometric complexity one encounters working with general closed paths. In particular we will assume the following restricted form of the well-known Jordan curve theorem which asserts that a simple closed path has an "inside" and an "outside." Suppose $\gamma$ is a curve with parameter interval $[\alpha, \beta]$, then by $\gamma^*$ we mean the image set of $\gamma$, in other words $\gamma^* := \{\gamma(t) : \ t \in [\alpha, \beta]\}$.

**Theorem 134** (Jordan curve theorem for a contour)**.** *Let $\gamma$ be a contour. Suppose that the inside of $\gamma$ is denoted by $I(\gamma)$ and the outside of $\gamma$ is denoted by $O(\gamma)$. Then the complement of $\gamma^*$ is of the form*

$$I(\gamma) \cup O(\gamma) \quad \textit{with} \quad I(\gamma) \cap O(\gamma) = \emptyset.$$

*Moreover both are open, connected sets with $I(\gamma)$ bounded and $O(\gamma)$ unbounded.*

For proof of Jordan's curve theorem, consult books on algebraic topology, such as [25].

**Remark 74.** Cauchy's theorem is often proved for a triangle first. For this version, suppose that $f$ is holomorphic inside and on a triangle $\gamma$. Then $\int_\gamma f(z)\,dz = 0$. There is also topologically more sophisticated approaches to Cauchy's theorem. For example, Cauchy's theorem for a contour assumes that $f$ is holomorphic inside and on a contour $\gamma$, then $\int_\gamma f(z)\,dz = 0$.

## Deformation Theorem

It is important to be able to study functions which are not holomorphic on the entire $\gamma$ and whose integral might not be zero. For example, $\frac{1}{z}$ fails to be holomorphic at $z = 0$ and

$$\int_\gamma f(z)dz = 2\pi i,$$

where $\gamma$ is the unit circle (the point $z = 0$ is called the singularity of $f$). To study such functions, we replace $\int_\gamma f(z)dz$ by $\int_{\widetilde{\gamma}} f(z)dz$ where $\widetilde{\gamma}$ is a simpler curve (say, a circle) where $\int_{\widetilde{\gamma}} f(z)dz$ can often be evaluated. The procedure that allows us to pass from $\gamma$ to $\widetilde{\gamma}$ is based on Cauchy's Theorem and is as follows. We first need the following definition.

## Preliminary Version of the Deformation Theorem

Let $f$ be holomorphic in a region $G$ and let $\gamma$ be a simple closed curve in $G$. Suppose that $\gamma$ can be continuously deformed to another simply closed curved $\widetilde{\gamma}$ without passing outside of $G$. (In such a case, we say that $\gamma$ is *homotopic* to $\widetilde{\gamma}$ in $G$). Then

$$\int_\gamma f(z)dz = \int_{\widetilde{\gamma}} f(z)\,dz. \tag{3.2}$$

Another version of the theorem is as follows: if $\gamma$ is a positively oriented contour, $\overline{D}(a;r) \subset I(\gamma)$ and $f$ is holomorphic inside $\gamma$ except possibly at $a$, then

$$\int_\gamma f(z)dz = \int_{\gamma(a;r)} f(z)\,dz.$$

Note that with this version we can replace an integral around a contour $\gamma$ by an integral around a circle, centered at a, inside $\gamma$.

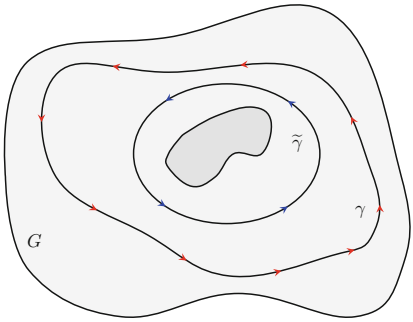



Figure 3.31: Curves $\gamma$ and $\widetilde{\gamma}$ are positively oriented

Note that $f$ need not be analytic inside $\gamma$, so that Cauchy's Theorem does not imply that the integrals in equation (3.2) are zero. It is implicit in the statement of the deformation theorem that both $\gamma$ and $\widetilde{\gamma}$ are oriented in the counterclockwise direction (Figure 3.31).

*Proof.* The strategy is to form a closed path to which Cauchy's theorem can be applied. Consider the following figure, in which a curve $\gamma_0$ is drawn joining $\gamma$ to $\widetilde{\gamma}$. We assume that such a curve can in fact be drawn (Figure 3.32).

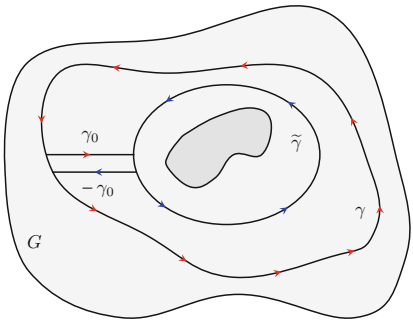



Figure 3.32: Preliminary version of the deformation theorem

We get a new curve consisting of $\gamma$, then $\gamma_0$, then $-\widetilde{\gamma}$, and then $-\gamma_0$ in that order. The inside of the curve is the shaded region. In this region, $f$ is holomorphic so we can apply Cauchy's Theorem:

$$\int_{\gamma+\gamma_0-\widetilde{\gamma}-\gamma_0} f = 0. \tag{3.3}$$

Strictly speaking this curve is not a simple closed curve, but such an objection can be taken care of by drawing two parallel copies of $\gamma_0$ and taking the limit as these copies converge together. From (3.3) we get

$$\int_{\gamma} f + \int_{\gamma_0} f - \int_{\widetilde{\gamma}} f - \int_{\gamma_0} f = 0 \implies \int_{\gamma} f = \int_{\widetilde{\gamma}} f$$

as desired. □

## Extension of Deformation Theorem

Suppose that $\gamma_1, \gamma_2, \ldots, \gamma_n$ are simply closed curves and that $\gamma$ is a simple closed curve with $f$ holomorphic on the region between $\gamma$ and $\gamma_1, \ldots, \gamma_n$. Then

$$\int_{\gamma} f = \sum_{k=1}^{n} \int_{\gamma_k} f.$$

*Proof.* Draw curves $\widetilde{\gamma}_1, \widetilde{\gamma}_2, \ldots, \widetilde{\gamma}_n$ joining $\gamma$ to $\gamma_1, \ldots, \gamma_n$, respectively, as shown in Figure 3.33.

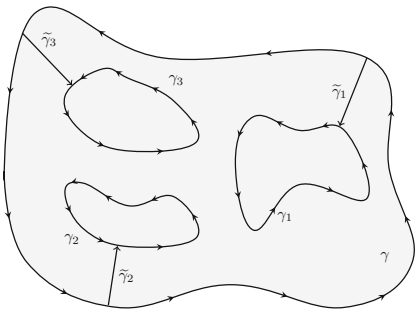



Figure 3.33: Extension of deformation theorem for $n = 3$

Let $\rho$ be the curve obtained by joining paths $\gamma, -\widetilde{\gamma}_1, -\gamma_1, \widetilde{\gamma}_1, \ \cdots \ , -\widetilde{\gamma}_n, -\gamma_n, \widetilde{\gamma}_n$ Since $f$ is holomorphic inside $\rho$,

$$\int_{\rho} f = 0.$$

But $\rho$ consists of $\gamma, -\gamma_1, -\gamma_2, \ldots, -\gamma_n$ and each $\widetilde{\gamma}_i$ traversed in opposite directions, so the contributions from this last portion cancel. Thus

$$0 = \int_{\gamma} f + \int_{-\gamma_1} f + \cdots + \int_{-\gamma_n} f = \int_{\gamma} f - \sum_{i=1}^{n} \int_{\gamma_i} f$$

as required. □

**Example 107.** Suppose we denote the integral of $f$ by $I = \int_{\gamma} f(z)dz$, where $\gamma(t) = e^{it}$ for $0 \leq t \leq 2\pi$ (the unit circle). For all of the following functions, we can show that $I = 0$.

a) Let $f(z) = \dfrac{e^{iz^2}}{4+z^2}$. This function is holomorphic except at $\pm 2i$, but these points are outside $\gamma(t)$. By Cauchy's Theorem, $I = 0$.

b) Let $f(z) = (\mathrm{Im} z)^2$. Note that $f(z)$ is not differentiable anywhere (use the Cauchy Riemann equations to show this). Thus we cannot apply Cauchy's Theorem. However, we can use the parametrization

$$\int_\gamma f(z)dz = \int_\gamma f(\gamma(t)) \cdot \gamma'(t)dt,$$

which results in

$$\begin{aligned} I = \int_0^{2\pi} (\sin t)^2 ie^{it} dt &= \int_0^{2\pi} (2\sin t \cos t) i(\cos t + i \sin t) dt \\ &= 2i \int_0^{2\pi} \sin t \cos^2 t dt - 2 \int_0^{2\pi} \sin^2 t \cos t dt = 0. \end{aligned}$$

c) Let $f(z) = \dfrac{1}{2z-1} - \dfrac{1}{2z+1}$. Observe that $f$ is not holomorphic at $z = \dfrac{1}{2}$ and $z = -\dfrac{1}{2}$ and both are contained in $\gamma(t)$, so we cannot use Cauchy's Theorem. In addition, the Fundamental Theorem of Calculus is not applicable . Can you think of a function $F(z)$ such that $F'(z) = f(z)$ in this region? But we can use the deformation theorem. Consider the following Figure 3.34.

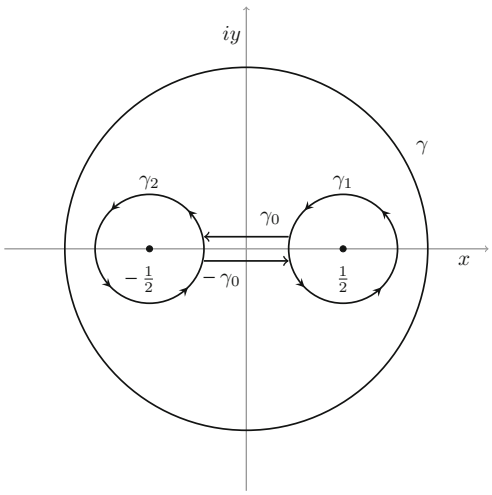



Figure 3.34: Illustrating example part c)

$$\text{Let} \quad \gamma_1(t) = \gamma\left(\frac{1}{2}; \frac{1}{4}\right) = \frac{1}{2} + \frac{1}{4}e^{it} \quad 0 \le t \le 2\pi \quad \text{and}$$
$$\gamma_2(t) = \gamma\left(-\frac{1}{2}; \frac{1}{4}\right) = -\frac{1}{2} + \frac{1}{4}e^{it} \quad 0 \le t \le 2\pi.$$

Let $\gamma_0$ be the straight line joining $\frac{1}{2}$ to $-\frac{1}{2}$. In addition, let $\widetilde{\gamma} : \gamma_1 + \gamma_0 + \gamma_2 - \gamma_0$. Then by the deformation theorem

$$\int_{\gamma} f(z)dz = \int_{\widetilde{\gamma}} f(z)dz = \int_{\gamma_1} f(z)dz + \int_{\gamma_2} f(z)dz.$$

Using the above parametrization, we get that

$$\begin{aligned}\int_{\gamma} f(z)dz &= \int_{\gamma_1} \left(\frac{1}{2z-1} - \frac{1}{2z+1}\right)dz + \int_{\gamma_2} \left(\frac{1}{2z-1} - \frac{1}{2z+1}\right)dz \\ &= \int_{\gamma_1} \frac{1}{2z-1}dz + \int_{\gamma_2} \frac{1}{2z+1}dz = 2\pi i \cdot \frac{1}{2} - 2\pi i \cdot \frac{1}{2} = 0.\end{aligned}$$

## Logarithms Again

Next we claim that there does not exist a holomorphic function $F$ defined on $\mathbb{C} \setminus \{0\}$ such that $F'(z) = \frac{1}{z}$. If such an $F$ existed, then using the Fundamental Theorem of Calculus, we could conclude that

$$\int_{\gamma} \frac{1}{z}dz = 0,$$

where $\gamma$ is a unit circle. However, we know that

$$\int_{\gamma} \frac{1}{z}dz = \int_0^{2\pi} \frac{ie^{it}}{e^{it}}dt = 2\pi i.$$

Therefore no such $F$ can exist. Although $\frac{d}{dz}(\log z) = \frac{1}{z}$, it does not contradict this example because $\log z$ is not holomorphic on $\mathbb{C} \setminus \{0\}$. It is holomorphic only on $\mathbb{C}$ minus the negative $x$-axis (including the origin) (Figure 3.35).

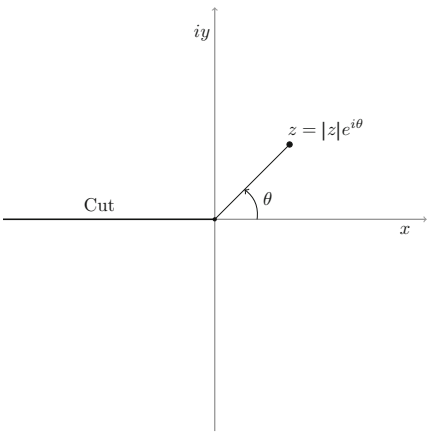



Figure 3.35: Cutting the negative $x$-axis including 0

Recall that $\log z : \mathbb{C} \setminus \{0\} \to \mathbb{C}$ is an inverse of $e^z$ when $e^z$ is restricted to the period strip $\{x + iy : y_0 \leq y \leq y_0 + 2\pi\}$. However, for differentiability of $\log z$ we must restrict $\log z$ to a set that is smaller than $\mathbb{C} \setminus \{0\}$. The reason is simple:

$$\log z = \log |z| + i \arg z, \quad \text{for say } 0 \leq \arg z \leq 2\pi.$$

But $\arg z$ is discontinuous. It jumps by $2\pi$ as we cross the real axis. If we remove the real axis, then we are excluding the usual positive reals on which we want $\log z$ defined. Therefore it is convenient to use the branch $-\pi < \arg z \leq \pi$ and an appropriate set on which $\log z$ is holomorphic.

**Proposition 43.** *Let $G$ be an open set that is given by*

$$G = \mathbb{C} \setminus \{x + iy : x \leq 0 \text{ and } y = 0\}.$$

*Define a branch of $\log z$ on $G$ by*

$$\log z = \log |z| + i \arg z \quad -\pi < \arg z \leq \pi,$$

*which is called the principal branch of the logarithm. Then $\log z$ is holomorphic on $G$ and $\dfrac{d}{dz} \log z = \dfrac{1}{z}$. Similar statements holds true for other branches as well.*

The next example shows how Cauchy's theorem can be used to solve real indefinite integrals.

**Example** We can use Cauchy's Theorem to show that

$$\int_0^\infty \frac{\sin^2 x}{x} dx = \frac{\pi}{4}.$$

We accomplish this by integrating $\dfrac{e^{iz^2}}{z}$ along the boundary of $r \le |z| \le R$ where $0 \le \arg z \le \pi/2$. Let $\gamma$ be the path obtained by joining $[r, R]$ to $\gamma_1$ to $-i[r, R]$ to $-\gamma_2$ (Figure 3.36). By Cauchy's Theorem,

$$\int_{\gamma} \frac{e^{iz^2}}{z} dz = 0.$$

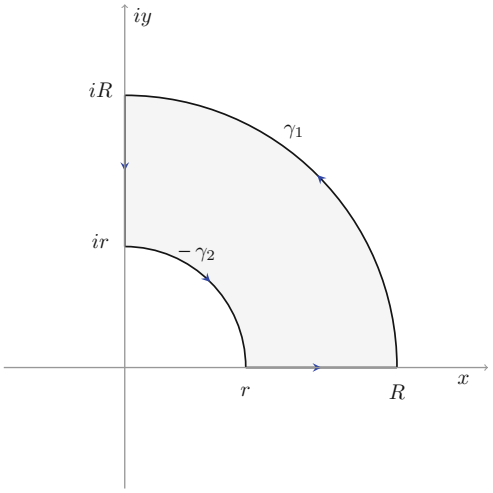



Figure 3.36: Contour $\gamma$

However, based on our construction of $\gamma$, we also have that

$$\int_{\gamma} \frac{e^{iz^2}}{z} dz = \int_{[r,R]} \frac{e^{iz^2}}{z} dz + \int_{\gamma_1} \frac{e^{iz^2}}{z} dz + \int_{-i[r,R]} \frac{e^{iz^2}}{z} dz + \int_{-\gamma_2} \frac{e^{iz^2}}{z} dz = 0. \quad (3.4)$$

We need to find all four integrals by considering the parametrization of each path.

**Integral along** $[r, R]$: If we write $z = x + iy$, then along $[r, R]$ we have that $y = 0$ so that $z = x$. Therefore,

$$\int_{[r,R]} \frac{e^{iz^2}}{z} dz = \int_r^R \frac{e^{ix^2}}{x} dx.$$

**Integral along** $\gamma_1$: On $\gamma_1$, $z = \gamma_1(\theta) = Re^{i\theta}$. Therefore

$$dz = Rie^{i\theta} d\theta, \quad e^{z^2} = e^{(Re^{i\theta})^2}.$$

Thus we have that

$$\int_{\gamma_1} \frac{e^{iz^2}}{z} dz = \int_0^{\pi/2} \frac{e^{i(Re^{i\theta})^2}}{Re^{i\theta}} Rie^{i\theta} d\theta = i \int_0^{\pi/2} e^{i(Re^{i\theta})^2} d\theta.$$

**Integral along** $-i[r, R]$: On $-i[r, R]$, we have that $z = 0 + iy$. Therefore, $dz = idy$ and

$$\frac{e^{iz^2}}{z} = \frac{e^{i(iy)^2}}{iy} = \frac{e^{-iy^2}}{iy}.$$

Thus

$$\int_{-i[r,R]} \frac{e^{iz^2}}{z} dz = -\int_{i[r,R]} \frac{e^{iz^2}}{z} dz = -\int_r^R \frac{e^{-iy^2}}{iy} idy = -\int_r^R \frac{e^{-iy^2}}{y} dy.$$

**Integral along** $-\gamma_2$: On $\gamma_2$, we have that $z = \gamma(\theta) = re^{i\theta}$. This implies that $dz = rie^{i\theta} d\theta$ and $e^{iz^2} = e^{i(re^{i\theta})^2}$. Therefore we have

$$\int_{-\gamma_2} \frac{e^{iz^2}}{z} dz = -\int_{\gamma_2} \frac{e^{iz^2}}{z} dz = -\int_0^{\pi/2} \frac{e^{i(re^{i\theta})^2}}{re^{i\theta}} rie^{i\theta} d\theta = -i \int_0^{\pi/2} e^{i(re^{i\theta})^2} d\theta.$$

Thus rewriting equation (3.4), we have that

$$\int_r^R \frac{e^{ix^2}}{x} dx + i \int_0^{\pi/2} e^{i(Re^{i\theta})^2} d\theta - \int_r^R \frac{e^{-iy^2}}{y} dy - i \int_0^{\pi/2} e^{i(re^{i\theta})^2} d\theta = 0.$$

The first and third above integrals can be combined as

$$\int_r^R \frac{e^{ix^2} - e^{-ix^2}}{x} dx.$$

Recalling the identity $\sin(x^2) = \dfrac{e^{ix^2} - e^{-ix^2}}{2i}$, the integral becomes

$$2i \int_r^R \frac{\sin(x^2)}{x} dx.$$

Since our problem is to evaluate $\displaystyle\int_0^\infty \frac{\sin(x^2)}{x} dx$, we must take $R \to \infty$ and $r \to 0$ and get estimates for the 2nd and 4th integrals.

**Claim #1:** The second integral in (3.4) vanishes as $R \to \infty$.
Proof: Observe that

$$\left| i \int_0^{\pi/2} e^{i(Re^{i\theta})^2} d\theta \right| \leq \int_0^{\pi/2} \left| e^{i(Re^{i\theta})^2} \right| d\theta. \tag{3.5}$$

However,

$$\left|e^{i(Re^{i\theta})^2}\right| = \left|e^{iR^2e^{2i\theta}}\right| = \left|e^{i[R^2(\cos(2\theta)+i\sin(2\theta))]}\right| = e^{-R^2\sin(2\theta)}.$$

(Note: recall that $|e^{iz}| = |e^{i(x+iy)}| = |e^{ix}\cdot e^{-y}| = |e^{ix}|e^{-y} = e^{-y}$). Going back to (3.5), we get

$$\left|i\int_0^{\pi/2} e^{i(Re^{i\theta})^2}d\theta\right| \leq \int_0^{\pi/2} e^{-R^2\sin(2\theta)}d\theta = 2\int_0^{\pi/4} e^{-R^2\sin(2\theta)}d\theta.$$

Think of $\sin(2\theta)$. The graph looks like Figure 3.37.

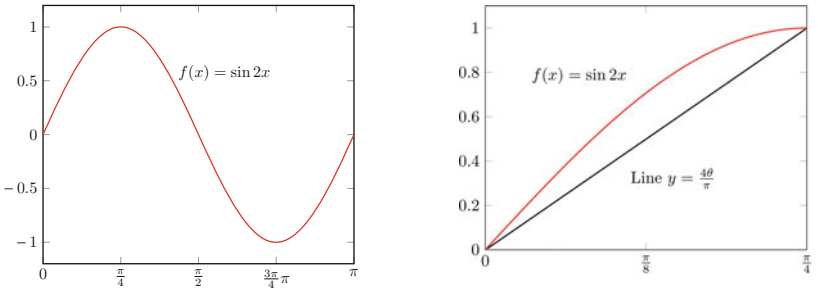



Figure 3.37: $\sin(2\theta) \geq \dfrac{4\theta}{\pi}$

When $\theta \in \left[0, \dfrac{\pi}{4}\right]$, we have that $\sin(2\theta) \geq \dfrac{4\theta}{\pi}$.

Note: one can also see $\sin(2\theta) \geq \dfrac{4\theta}{\pi}$ by setting up a function $h(\theta) = \sin(2\theta) - \dfrac{4\theta}{\pi}$ and observing that $h(0) = 0$ and $h(\frac{\pi}{4}) = 0$ while $h''(0) < 0$ for $0 < \theta < \dfrac{\pi}{4}$, implying that $\sin(2\theta) \geq \dfrac{4\theta}{\pi}$.

Continuing with (3.5), we see that

$$2\int_0^{\pi/4} e^{-R^2\sin(2\theta)}d\theta < 2\int_0^{\pi/4} e^{-R^2\frac{4\theta}{\pi}}d\theta.$$

We can integrate this last integral. In fact,

$$2\int_0^{\pi/4} e^{-\frac{4R^2}{\pi}\theta}d\theta = -2\frac{\pi}{4R^2}e^{-\frac{4R^2}{\pi}\theta}\Big|_0^{\pi/4} = -\frac{\pi}{2R^2}[e^{-R^2} - 1] = \frac{\pi}{2R^2}[1 - e^{-R^2}].$$

Now as $R \to \infty$, the second integral in (3.4) will therefore vanish.

**Claim #2:** The fourth integral is equal to $i\dfrac{\pi}{2}$ as $r \to 0$.

Proof: We use the continuity of $e^{iz^2}$ at $z = 0$. When $z = 0$, $e^{iz^2} = 1$. Given any $\epsilon > 0$ there must exist a $r > 0$ such that

$$|e^{iz^2} - 1| < \epsilon$$

whenever $|z| < r$. Therefore

$$\left| i \int_0^{\pi/2} e^{i(re^{i\theta})^2} d\theta - i\pi/2 \right| = \left| i \int_0^{\pi/2} \left( e^{i(re^{i\theta})^2} - 1 \right) d\theta \right| < \epsilon \cdot \frac{\pi}{2}.$$

Thus the th integral approaches $i\frac{\pi}{2}$ as $r \to 0$.

Going back to equation (3.4), combining the st and rd integrals and using Claim #1 and Claim #2, we have

$$\int_0^\infty \frac{e^{ix^2} - e^{-ix^2}}{x} dx - i\frac{\pi}{2} = 2i \int_0^\infty \frac{\sin(x^2)}{x} dx - i\frac{\pi}{2} = 0.$$

Thus

$$\int_0^\infty \frac{\sin(x^2)}{x} dx = \frac{\pi}{4}$$

as desired.

## Exercises

1. Evaluate the following integrals:

   a) $\int_\gamma e^z \, dz$ where $\gamma$ is the perimeter of the unit square

   b) $\int_\gamma \frac{1}{z^2} \, dz$ where $\gamma$ is the unit circle

   c) $\int_\gamma \frac{1}{z} \, dz$ where $\gamma$ is the circle $3 + e^{i\theta}$, $0 \le \theta \le 2\pi$

   d) $\int_\gamma z^2 dz$ where $\gamma$ is the segment joining $1 + i$ to 2

2. For what simple closed curves $\gamma$ does the following hold:

$$\int_\gamma \frac{dz}{z^2 + z + 1} = 0.$$

3. Does Cauchy's theorem hold separately for real and imaginary parts of $f$? If so, prove that it does; if not, give a counterexample.

4. Let $\gamma$ be a piecewise smooth closed curve. Show that

$$\int_\gamma z^m dz = 0, \qquad m = 0, 1, 2, \ldots$$

5. Use Green's theorem (complex form) to show that

$$\frac{1}{2\pi i} \int_\gamma \frac{dz}{z - p} = \begin{cases} 0 & \text{if } p \text{ is outside } \gamma \\ 1 & \text{if } p \text{ is inside } \gamma. \end{cases}$$

6. Let $D$ be an arbitrary region and $R = \{x + iy : a \le x \le b, \quad c \le y \le d\}$ be a rectangle. Show that if $R \subseteq D$ and $f$ is differentiable in $D$, then

$$\int_{\partial R} f = 0,$$

where $\partial R = [z_1, z_2] + [z_2, z_3] + [z_3, z_4] + [z_4, z_1]$.

7. Let $f(z)$ be holomorphic on a simple connected region $A$, except possibly not holomorphic at $z_0 \in A$. Suppose, however, that $f$ is bounded in absolute value near $z_0$. Show that for any simple closed curve $\gamma$ containing $z_0$,

$$\int_\gamma f(z)\, dz = 0.$$

8. Let $f$ be holomorphic on $\mathbb{C}$ (such functions are called entire functions). Evaluate

$$\int_0^{2\pi} f(z_0 + re^{i\theta})e^{i\theta} d\theta.$$

9. Let $f$ be a continuous function in the disc $|z - z_0| < r$ and let $\gamma_\epsilon$ be the circle $|z - z_0| = \epsilon$. Show that

$$\lim_{\epsilon \to 0} \frac{1}{2\pi i} \int_{\gamma_\epsilon} \frac{f(z)}{z - z_0}\, dz = f(z_0).$$

10. Let $f(z)$ be a holomorphic function defined for $|z| \le 1$ and $u(x, y) = \mathrm{Re} f(z)$, where $z = x + iy$. Show that

$$\int_\gamma \frac{\partial u}{\partial y} dx - \frac{\partial u}{\partial x} dy = 0,$$

where $\gamma$ is the unit circle.

11. Show that

$$\int_0^\infty \cos x^2\, dx = \int_0^\infty \sin x^2\, dx = \sqrt{\pi/8}.$$

Hint: Integrate $e^{iz^2}$ round the contour $\gamma$ shown in Figure 3.38. Use Cauchy's theorem and the inequality $\sin 2\theta \ge \dfrac{4\theta}{\pi}$ for $\theta \in \left[0, \dfrac{\pi}{4}\right]$.

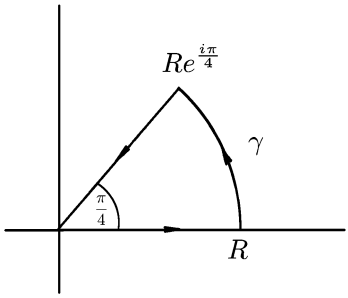



Figure 3.38: Contour for Exercise 11

## 3.8 Cauchy's Formulae

### Cauchy's Integral Formula

**Theorem 135.** *Let $f$ be holomorphic inside and on a positively oriented contour $\gamma$. Then if $a$ is inside $\gamma$,*

$$f(a) = \frac{1}{2\pi i} \int_\gamma \frac{f(w)}{w-a}\, dw.$$

The above formula for $f(a)$ is remarkable in the sense that it states the values of $f$ on $\gamma$ completely determine the values of $f$ inside $\gamma$. In other words, the value of $f$ is determined by its "boundary values."

*Proof.* First note that since $f(a)$ is constant, we have

$$\int_\gamma \frac{f(a)}{w-a}\, dw = f(a) \int_\gamma \frac{1}{w-a}\, dw = 2\pi i f(a)$$

from the fundamental integral. We also know by the Jordan Curve Theorem that there exists $r$ such that $D(a;r)$ lies inside $\gamma$. For any $\epsilon < r$,

$$\int_\gamma \frac{f(w)}{w-a}\, dw = \int_{\gamma(a;\epsilon)} \frac{f(w)}{w-a}\, dw,$$

by the deformation theorem. Now consider the difference

$$\left| \frac{1}{2\pi i} \int_\gamma \frac{f(w)}{w-a}\, dw - f(a) \right| = \left| \frac{1}{2\pi i} \int_{\gamma(a;\epsilon)} \frac{f(w)-f(a)}{w-a}\, dw \right|$$

and use the estimation lemma to obtain

$$\left| \frac{1}{2\pi i} \int_\gamma \frac{f(w)}{w-a}\, dw - f(a) \right| \leq \frac{1}{2\pi}\, 2\pi \sup_{t\in[0,2\pi]} |f(a+\epsilon e^{it}) - f(a)|.$$

Now all we have to do is to remember that $f$ is continuous at $a$, so the supremum above tends to zero as $\epsilon \to 0$. Thus the left-hand side must be zero. □

**Example 108.**

a) Suppose $\gamma$ is the unit circle $|z| = 1$, then using Cauchy's integral formula:

$$\int_\gamma \frac{e^z + \sin z}{z}\, dz = 2\pi i f(0) = 2\pi i.$$

Here the denominator of the integrand is $z = z - 0$ and $f(w) = e^w + \sin w$, which is holomorphic inside and on $\gamma$. The value 0 is in $\gamma$, so we can apply Cauchy's Integral Theorem with $f(0) = 1$ to obtain the claim.

b) Evaluating

$$\int_\gamma \frac{z^2}{z^2+1}\, dz,$$

where $\gamma = \gamma(i, 1)$, we first observe $z^2 + 1 = (z + i)(z - i) = 0$. For $z = i$ and $z = -i$, $\dfrac{z^2}{z^2+1}$ is not holomorphic, but only $z = i$ is inside $\gamma$. By rewriting

$$\int_\gamma \frac{z^2}{z^2+1}\, dz = \int_\gamma \frac{\dfrac{z^2}{z+i}}{z-i}\, dx = 2\pi i f(i) = -\pi.$$

Because $f(w) = \dfrac{w^2}{w+i}$, which is holomorphic inside and on $\gamma$, by substituting $f(i) = \dfrac{i^2}{i+i} = -\dfrac{1}{2i}$ into Cauchy's integral formula we prove the claim.

## Liouville's Theorem

**Theorem 136** (Liouville)**.** *Let $f$ be holomorphic and bounded in the complex plane $\mathbb{C}$. Then $f$ is constant.*

*Proof.* Suppose $|f(w)| \le M$ for all $w \in \mathbb{C}$. Fix $a, b$ in $\mathbb{C}$, and take $R$ such that

$$|w - a| \ge \frac{1}{2}R \quad \text{and} \quad |w - b| \ge \frac{1}{2}R \quad \text{whenever} \quad |w| = R.$$

Cauchy's integral formula applied to $f(a)$ and $f(b)$ with $\gamma(0, R)$ and then applying the estimation lemma yields

$$|f(a) - f(b)| \le \frac{1}{2\pi}\,(2\pi R) M \frac{|a - b|}{(1/2R)^2}.$$

Thus $f(a) = f(b)$ when $R$ is sufficiently large. □

**Remark 75.** In real analysis there are many infinitely differentiable functions that are bounded, i.e., $|f(x)| \leq M$ for all $x \in \mathbb{R}$. For example, $f(x) = \sin x$ is infinitely differentiable and $|\sin x| \leq 1$ for all $x$, but it is **not** a constant function. In complex analysis, Liouville's theorem asserts that the only bounded entire functions (entire means holomorphic in the whole complex plane $\mathbb{C}$) are constant functions.

**Example 109.** Let $M \geq 0$ and $f$ be an entire function such that $\operatorname{Re} f(z) \leq M$ for all $z \in \mathbb{C}$. Show that $f$ is a constant function.
Consider the function

$$F(z) = e^{f(z)} \quad \text{for all} \quad z \in \mathbb{C}.$$

Then, $F(z)$ is an entire function and $|F(z)| = |e^{f(z)}| = e^{\operatorname{Re} f(z)}$ and thus

$$|F(z)| \leq e^M.$$

By Liouville's theorem $F(z)$ is constant and therefore $f$ must be a constant function.

## Fundamental Theorem of Algebra

**Theorem 137.** *Let $p(z)$ be a non-constant polynomial with complex coefficients. Then, there exists a complex number $\xi$ such that $p(\xi) = 0$. Consequently a complex polynomial of degree $n > 1$ has $n$ roots (not necessarily distinct) in $\mathbb{C}$.*

*Proof.* The proof is done by contradiction. Suppose we have a non-constant polynomial $p(z)$ and $p(z) \neq 0$ for every $z \in \mathbb{C}$. Then, $\dfrac{1}{p(z)}$ is holomorphic. Next, we claim that $\dfrac{1}{p(z)}$ is bounded on $\mathbb{C}$. Since $|p(z)| \to \infty$ as $|z| \to \infty$, there exists $R$ such that

$$\left|\frac{1}{p(z)}\right| < 1 \quad \text{for} \quad |z| > R.$$

On the other hand, on the compact set $|z| \leq R$, the function $\dfrac{1}{p(z)}$ is continuous and hence bounded (it is bounded because it is a continuous functions on a compact set). By Liouville's theorem $\dfrac{1}{p(z)}$ must be constant, a contradiction. The last statement in the theorem is proved by induction on $n$ (the degree of the polynomial). □

**Example 110.** Show that every polynomial $p(z) = a_n z^n + \cdots + a_0$ of degree $n \geq 1$ has precisely $n$ roots in $\mathbb{C}$. If these roots are denoted by $w_1, w_2, \ldots, w_n$ then $p$ can be factored as

$$p(z) = a_n(z - w_1)(z - w_2)\cdots(z - w_n).$$

By the fundamental theorem of algebra $p$ has a root, say $w_1$. Then writing $z = (z - w_1) + w_1$ and inserting this for $z$ in $p$, using the binomial formula we get

$$p(z) = b_n(z - w_1)^n + \cdots + b_1(z - w_1) + b_0,$$

where $b_0, b_1, \ldots, b_n$ are new coefficients and $b_n = a_n$. Since $p(w_1) = 0$ we know $b_0 = 0$, therefore we can factor out $(z - w_1)$ and write

$$p(z) = (z - w_1)q(z),$$

where $q(z)$ is a polynomial of degree $n - 1$. By induction on the degree of the polynomial we conclude that

$$p(z) = c(z - w_1)(z - w_2)\cdots(z - w_n)$$

for some $c \in \mathbb{C}$. Expanding the right-hand side shows that the coefficient of $z^n$ is $c$ and therefore $c = a_n$ as desired.

## Morera's Theorem

Morera's theorem gives a sufficient condition for $f$ being holomorphic and is often taken as the converse of Cauchy's theorem. There are several versions of this theorem. For proofs of the versions below we refer the reader to [41], p. 156.

**Theorem 138.** *Version 1: Suppose $f$ is continuous on an open set $G$ and that $\int_\gamma f(z)\,dz = 0$ for all* **triangles** *$\gamma$ in $G$. Then $f$ is holomorphic in $G$.*
*Version 2: Suppose $f$ is continuous in a simply connected region $G$ and that $\int_\gamma f(z)\,dz = 0$ for all* **closed contour** *$\gamma$. Then $f$ is holomorphic in $G$.*

## Cauchy's Formula for Derivatives and its Consequences

**Theorem 139.** *Let $f$ be holomorphic inside and on a positively oriented contour $\gamma$ and let $a$ lie inside $\gamma$. Then $f^n(a)$ exists for $n = 1, 2, \ldots$ and*

$$f^n(a) = \frac{n!}{2\pi i}\int_\gamma \frac{f(w)}{(w - a)^{n+1}}\,dw.$$

The proof of this theorem is in most complex analysis books, for example consult [41], p. 157. We observe that if we can interchange $\int$ and the $\frac{d}{dz}$, one can obtain an expression for $f'(z)$ by taking the derivative under the integral sign in Cauchy's integral formula $f(z) = \frac{1}{2\pi i}\int_\gamma \frac{f(w)}{w - z}\,dw$;

$$f'(z) = \frac{d}{dz}f(z) = \frac{d}{dz}\left(\frac{1}{2\pi i}\int_\gamma \frac{f(w)}{w - z}\right) = \frac{1}{2\pi i}\int_\gamma \frac{d}{dz}\left(\frac{f(w)}{w - z}\right)dw$$

and therefore

$$f'(z) = \frac{1}{2\pi i}\int_\gamma \frac{f(w)}{(w-z)^2}dw.$$

We may repeat this procedure by differentiating $\frac{1}{2\pi i}\int_\gamma \frac{f(w)}{(w-z)^2}dw$ with respect to $w$, and hope to obtain $f''(z) = \frac{2}{2\pi i}\int_\gamma \frac{f(w)}{(w-z)^3}dw$ where $z \in I(\gamma)$. There is a difference between this second differentiation and the first one. Since $f$ is holomorphic $f'(z)$ exists, but we do not know that $f''(z)$ exists. However we can use Cauchy's integral formula to examine the quotient $\frac{f'(z+h)-f'(z)}{h}$ and then show that as $h \to 0$ this limit exists and it is given by the right-hand side of $f'(z)$ above. Then the process can be repeated to show successively, the existence of $f^n(z)$ for $n = 3, 4, \ldots$ all by differentiating under the integral sign to obtain

$$f^n(a) = \frac{n!}{2\pi i}\int_\gamma \frac{f(w)}{(w-a)^{n+1}}\ dw.$$

**Example 111.** Suppose we want to evaluate:

$$\int_\gamma \frac{z+1}{z^4+2iz^3}\ dz$$

where $\gamma : |z| = 1$.

Since $z^4+2iz^3 = z^3(z+2i)$, $\frac{z+1}{z^4+2iz^3}$ is not holomorphic at $z = 0$ and $z = -2i$. However only $z = 0$ lies inside the contour $\gamma$. By rewriting:

$$\frac{z+1}{z^4+2iz^3} = \frac{\frac{z+1}{z+2i}}{z^3}$$

we can identify $a = 0$, $n = 2$ and $f(z) = \frac{z+1}{z+2i}$. The quotient rule for derivatives gives $f''(z) = \frac{2-4i}{(z+2i)^3}$ and so $f''(0) = \frac{2i-1}{4i}$. Hence the given integral

$$\int_\gamma \frac{z+1}{z^4+2iz^3}\ dz = \frac{2\pi i}{2!}f''(0) = -\frac{\pi}{4} + \frac{\pi}{2}i.$$

## Cauchy's Inequality

**Theorem 140.** *Let $f$ be holomorphic in a simply connected region $G$ and $\gamma$ be the simple closed curve $\gamma(a, R)$ which lies entirely in $G$. If $|f(z)| \le M$ for all points $z$ on $\gamma$, then*

$$|f^n(a)| \le \frac{n!M}{R^n} \qquad \textit{(Cauchy's Inequality).}$$

*Proof.* From Cauchy's integral formula for derivatives, we have

$$f^{(n)}(a) = \frac{n!}{2\pi i} \int_\gamma \frac{f(z)}{(z-a)^{n+1}}\, dz.$$

Applying the estimation lemma, for $\gamma$, the circle with radius $R$, centered at $a$, we obtain

$$\begin{aligned} \left|f^{(n)}(a)\right| &= \frac{n!}{2\pi} \left| \int_\gamma \frac{f(z)}{(z-a)^{n+1}}\, dz \right| \\ &\leq \frac{n!}{2\pi} \int_\gamma \frac{\left|f(z)\right|}{\left|z-a\right|^{n+1}} \left|dz\right| \\ &\leq \frac{n!}{2\pi} \frac{M}{R^{n+1}} 2\pi R = \frac{n!}{R^n} M. \quad \square \end{aligned}$$

## Taylor's Theorem

We have seen in Section 3.3, Theorem 125 that if $f(z) = \sum_{n=0}^{\infty} c_n z^n$ and if $\sum_{n=0}^{\infty} c_n z^n$ has a radius of convergence $R \neq 0$, then

- $f \in H(D(0;R))$.
- $f'(z) = \sum_{n=1}^{\infty} n c_n z^{n-1}$ when $|z| < R$.

This means that convergent complex power series define a holomorphic function. How about the converse of this theorem? The answer to this question is called Taylor's theorem. By using Cauchy's integral formula one can prove that any function holomorphic in a disc has a power series expansion.

**Theorem 141** (Taylor's theorem)**.** *Let $f$ be holomorphic in $D(a;R)$ with $R > 0$. Then there exists unique constants $c_n$ such that*

$$f(z) = \sum_{n=0}^{\infty} c_n (z-a)^n \quad z \in D(a;R),$$

*where the constant $c_n$ is given by the formula*

$$c_n = \frac{f^{(n)}(a)}{n!} = \frac{1}{2\pi i} \int_\gamma \frac{f(w)}{(w-a)^{n+1}}\, dw$$

*and $\gamma$ is a circle $\gamma(a;r)$ for $0 < r < R$.*

*Proof.* Fix a point $z \in D(a;R)$ and choose $r$ such that $|z-a| < r < R$. Using Cauchy's integral formula write

$$f(z) = \frac{1}{2\pi i} \int_\gamma \frac{f(w)}{w-z} dw.$$

Since $|z-a| < |w-a|$ for all $w$ on $\gamma$ as shown in Figure 3.39,

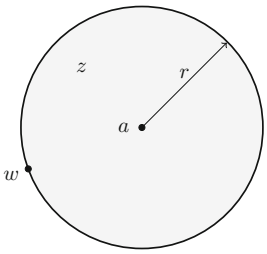



Figure 3.39: Proof of Taylor's theorem

$$\frac{1}{w-z} = \frac{1}{(w-a)\left(1-\frac{z-a}{w-a}\right)} = \sum_{n=0}^{\infty} \frac{(z-a)^n}{(w-a)^{n+1}}.$$

Therefore, Cauchy's integral formula above takes the form

$$f(z) = \frac{1}{2\pi i}\int_\gamma \frac{f(w)}{w-z}\ dw = \frac{1}{2\pi i}\int_\gamma \sum_{n=0}^{\infty} \frac{(z-a)^n}{(w-a)^{n+1}}\, f(w)\, dw.$$

Now the question is can we interchange summation and integration? The answer to this question is yes if one has uniform convergence. We appeal to the Weierstrass M-test to see that we can do this. Then

$$f(z) = \sum_{n=0}^{\infty} \left(\frac{1}{2\pi i}\int_\gamma \frac{f(w)}{(w-a)^{n+1}}\, dw\right)(z-a)^n.$$

Now we appeal to Cauchy's integral formula for derivatives to finish the proof. It is left to the reader to prove the uniqueness of $c_n$. □

There are a wealth of interesting results whose proofs rest on Cauchy's theorem and the Cauchy integral formulas. In the following we mention one more without proof. Recall that if $f$ is continuous on a closed and bounded region $R$, then $f$ is bounded; that is there is some constant $M$ such that $|f(z)| \leq M$ for all $z \in R$. If the boundary of $R$ is a simple closed curve $\gamma$, then the next theorem tells us that $|f(z)|$ assumes its maximum value at some point on the boundary of $\gamma$.

## Maximum Modulus Theorem

**Theorem 142** ( [5], p. 134)**.** *Suppose $f$ is holomorphic and non-constant on a closed simple curve $\gamma$. Then the modulus $|f(z)|$ attains its maximum on $\gamma$ (on the boundary).*

**Remark 76.**

a) If we assume $f(z) \neq 0$ in the closed region bounded by $\gamma$, then by the hypotheses of the maximum modulus theorem $|f(z)|$ also attains its minimum on $\gamma$.

b) In real analysis there are many infinitely differentiable functions that are bounded on a closed interval $[a, b]$ but do not attain their maximum at the boundary of $[a, b]$. For example $\sin x$ is infinitely differentiable and bounded on $[-\pi, \pi]$ and $|\sin(\pi)| = |\sin(-\pi)| = 0$ but $|\sin x| \leq 1$ for all $x$. The situation is different in complex analysis, where maximum modulus is attained on the boundary.

**Example 112.** Find the maximum of $|f(z)| = |2z + 5i|$ on $|z| \leq 2$.
Since $|z|^2 = z\overline{z}$, we have

$$\begin{aligned} |f(z)|^2 = |2z + 5i|^2 &= (2z + 5i)\overline{(2z + 5i)} = (2z + 5i)(2\overline{z} - 5i) \\ &= 4z\overline{z} - 10i(z - \overline{z}) + 25 \end{aligned}$$

Since $z - \overline{z} = 2i\,\mathrm{Im}(z)$, the above expression takes the form

$$|f(z)|^2 = |2z + 5i|^2 = 4|z|^2 + 20\mathrm{Im}(z) + 25.$$

The function $f(z)$ is a polynomial and thus holomorphic on the region defined by $|z| \leq 2$. By the maximum modulus theorem its maximum occurs on the boundary $|z| = 2$. Then from $|2z + 5i|^2 = 4|z|^2 + 20\mathrm{Im}(z) + 25$, and $|z| = 2$ we get

$$|2z + 5i| = \sqrt{16 + 20\mathrm{Im}(z) + 25}.$$

As $z = x + iy$, and $|\mathrm{Im}(z)| \leq |z| = 2$, $\mathrm{Im}(z)$ attains its maximum at the point $z = 2i$. Thus,

$$\max_{|z| \leq 2} |2z + 5i| = \sqrt{81} = 9.$$

## Schwarz' Lemma

The maximum modulus principle is a powerful tool for obtaining bounds for the size of holomorphic, non-constant functions. There is a simple but important consequence of the maximum modulus theorem called Schwarz' lemma.

**Lemma 17.** *Let $f$ be holomorphic in the unit disc $\mathbb{D}(0, 1)$, with $f(0) = 0$ and $|f(z)| < 1$ in $\mathbb{D}(0, 1)$, Then, $|f'(0)| \leq 1$ and $|f(z)| \leq |z|$ in $\mathbb{D}(0, 1)$. Strict inequality holds in both estimates unless $f$ is a rotation of the disc: $f(z) = e^{i\theta}z$.*

For the proof of the Schwarz Lemma, we refer the reader to [5], pp. 135–136. Ahlfors gave a generalization of the proof of Schwarz's lemma in terms of curvature of the metric [4]. This extended proof can be thought of as a starting point of an area in complex analysis called Geometric Function Theory (GFT). To learn more about GFT consult [34].

## Exercises

1. Show that

$$\int_\gamma \frac{\cos z}{z(z^2+1)}\,dz = \begin{cases} 2\pi i(1-\cos i) & \gamma : |z| = 3, \\ 2\pi i & \gamma : |z| = \dfrac{1}{3}, \\ 0 & \gamma : |z-1| = \dfrac{1}{3}. \end{cases}$$

2. Let $M \geq 0$ and let $f$ be an entire function such that $|f(z)| \leq M$ for all $z \in \mathbb{C}$. Show that $f$ is constant.

3. Show the following identities:

i. $$\int_{|z|=2} \frac{9z^2 - iz + 4}{z(z^2+1)}\,dz = 18\pi i,$$

ii. $$\int_\gamma \frac{e^{3z} + 3\cosh z}{\left(z - \dfrac{i\pi}{2}\right)^4}\,dz = 8\pi.$$

where $\gamma$ is a simple closed contour containing $\dfrac{i\pi}{2}$ in its interior.

4. Prove that every polynomial equation $p(z) = a_0 + a_1 z + \cdots + a_n z^n = 0$ with degree $n \geq 1$ and $a_n \neq 0$ has exactly $n$ roots.

5. Prove that if $f(z)$ is holomorphic inside and on a circle $\gamma$ with center $a$ and radius $r$, then $f(a)$ is the mean of the values of $f(z)$ on $\mathbb{C}$,

$$\text{i.e.,} \quad f(a) = \frac{1}{2\pi}\int_0^{2\pi} f(a + re^{it})\,dt.$$

(This is called *Gauss' Mean Value Theorem.*)

6. Suppose $f$ is holomorphic inside and on a positively oriented contour $\gamma$. Let $a$ lie inside $\gamma$. Show that $f'(a)$ exists and $f'(a) = \dfrac{1}{2\pi i}\displaystyle\int \frac{f(w)}{(w-a)^2}\,dw$.

(Hint: Use Cauchy's integral formula and show that

$$\left| \frac{f(a+h) - f(a)}{h} - \frac{1}{2\pi i}\int \frac{f(w)}{(w-a)^2}\,dw \right| \longrightarrow 0$$

as $h \longrightarrow 0$.)

7. Let $f(z)$ be entire and let $|f(z)| \geq 1$ on the whole complex plane. Prove that $f$ is constant.

8. Let $f$ be continuous on a region $A$ and holomorphic on $A \setminus \{z_0\}$ for a point $z_0 \in A$. Show that $f$ is holomorphic on $A$.

9. Show that if $F$ is holomorphic on $A$, then so is $f$ where $f(z) = \dfrac{F(z) - F(z_0)}{z - z_0}$, if $z \neq z_0$ and $f(z_0) = F'(z_0)$ where $z_0$ is some point in $A$.

10. Let $f$ be holomorphic on a region $A$ and let $\gamma$ be a circle with radius $R$ and center $z_0$ that lies in $A$. Assume that the disc $\{z : |z - z_0| < R\}$ also lies in $A$. Suppose that $|f(z)| \leq M$ for all $z \in \gamma$. Then show that for any $k = 0, 1, 2, \ldots,$

$$|f^{(k)}(z_0)| \leq \frac{k!}{R^k} M.$$

11. Find the maximum of $|\sin z|$ on $[0, 2\pi] \times [0, 2\pi]$.

12. Let $f(z) = z^2 + 4z - 1$ be defined on $|z| \leq 1$. Show that the maximum modulus of $f(z)$ over the region is 6.

13. If $f$ is holomorphic in a closed region G bounded by a simple closed curve $\gamma$ and $f(z) \neq 0$ for all $z \in G$, prove that the modulus $|f(z)|$ attains its minimum on $\gamma$.

14. Let the function $f$ be holomorphic in the entire complex plane and satisfy

$$\int_0^{2\pi} |f(re^{i\theta}| \, d\theta \leq r^{\frac{17}{3}}$$

for all $r > 0$. Prove that $f$ is a zero function.

## 3.9 Laurent Expansion and Singularities

### Uniformly Convergent Series and Weierstrass' M-test

Let $S$ be a nonempty subset of $\mathbb{C}$ and let $\{f_n\}$ be a sequence of complex-valued functions defined on $S$. We say $\{f_n\}$ converges uniformly to $f$ on $S$ if

$$d_n := \sup \Big\{ |f_n(z) - f(z)| : z \in S \Big\} \to 0 \quad \text{as } n \to \infty.$$

As usual, we say $\displaystyle\sum_{k=0}^{\infty} u_k(z)$ converges uniformly on $S$ if the sequence of partial sums $f_n = u_1 + u_2 + \cdots + u_n$ converges uniformly on $S$. Fortunately, there is one easy sufficient condition for uniform convergence of series, called the Weierstrass M-test (see Section 2.3).

**Theorem 143** (Weierstrass M-test)**.** *The series $\sum u_k$ converges uniformly on $S$ if there exist real numbers $M_k$ such that for all $k \in \mathbb{N}$*

$$|u_k(z)| \leq M_k \quad \textit{for all } z \in S \quad \textit{and} \quad \sum_{k=1}^{\infty} M_k \textit{ converges.}$$

The proof of this theorem is very similar to the one we gave in Chapter 2. We invoke the fact that $\mathbb{C}$ is complete, thus a sequence $\{z_n\}$ converges if and only if it is a Cauchy sequence.

The next theorem is about the relationship between the radius of convergence of a power series and uniform convergence.

**Theorem 144.** *Let $\sum c_k z^k$ be a complex power series with a radius of convergence $R > 0$. Then $\sum c_k z^k$ converges uniformly on any closed disc $\overline{D}(0; R - \epsilon)$ for any $\epsilon > 0$ (Figure 3.40).*

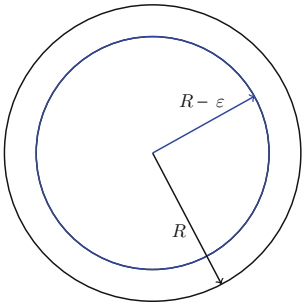



Figure 3.40: Pointwise convergence versus uniform convergence

*Proof.* Recall that if $\sum c_k z^k$ has a radius of convergence $R$, then $\sum c_k z^k$ converges absolutely for all $|z| < R$. Now observe that

$$|c_k z^k| \leq M_k := |c_k|(R - \epsilon)^k \quad \text{since } |z| \leq R - \epsilon < R.$$

Therefore $\sum M_k$ converges by the above fact and thus $\sum c_k z^k$ converges uniformly on any curve $\gamma = \gamma(0; r)$ where $0 \leq r < R$. □

**Remark 77.** Note that in general $\sum c_k z^k$ does not converge uniformly on $D(0; R)$. Consider the geometric series $\sum_{k=0}^{\infty} z^k$, where $R = 1$. For $|z| < 1$,

$$f_n(z) = 1 + z + z^2 + \cdots + z^n = \frac{1 - z^{n+1}}{1 - z}$$

which converges pointwise to $f(z) = \dfrac{1}{1-z}$. However, the convergence is not uniform. To see this, consider

$$d_n := \sup\left\{|f_n(z) - f(z)| : |z| < 1\right\} = \sup\left\{\left|\frac{1-z^{n+1}}{1-z} - \frac{1}{1-z}\right| : |z| < 1\right\}$$
$$= \sup\left\{\left|\frac{z^{n+1}}{1-z}\right| : |z| < 1\right\} \geq \sup\left\{\frac{x^{n+1}}{1-x} : 0 \leq x < 1\right\}.$$

One can see that $\dfrac{x^{n+1}}{1-x}$ is strictly increasing on $[0,1)$ and thus $\dfrac{x^n}{1-x} \to \infty$ as $x \to 1$. Thus $d_n \not\to 0$ and $\sum z^k$ does not converge uniformly on $D(0;1)$.

## Laurent Series

By a series of the form $\displaystyle\sum_{n=-\infty}^{\infty} c_n(z-a)^n$ we mean

$$\sum_{n=-\infty}^{\infty} c_n(z-a)^n = \sum_{n=1}^{\infty} c_{-n}(z-a)^{-n} + \sum_{n=0}^{\infty} c_n(z-a)^n$$
$$= \sum_{n=1}^{\infty} \frac{b_n}{(z-a)^n} + \sum_{n=0}^{\infty} c_n(z-a)^n$$

where we set $c_{-n} = b_n$. By the convergence of $\displaystyle\sum_{n=-\infty}^{\infty} c_n(z-a)^n$ we mean convergence of both series on the right. A power series converges inside a disc, and consequently a power series with negative powers alone should converge outside a disc. Those with positive and negative powers should therefore converge in a region between two concentric circles. Such a region is called an annulus.

**Theorem 145** (Laurent Series)**.** *Suppose $f$ is holomorphic on $A = \{z : R_1 < |z-a| < R_2\}$ where we allow $R_1 = 0$, $R_2 = \infty$, or both. Then there is a unique sequence $\{c_n\} \in \mathbb{C}$ such that*

$$f(z) = \sum_{n=-\infty}^{\infty} c_n(z-a)^n$$

*where both series converge absolutely on $A$. Moreover,*

$$c_n = \frac{1}{2\pi i}\int_\gamma \frac{f(w)}{(w-a)^{n+1}}dw$$

*where $\gamma = \gamma(a;R)$, oriented counterclockwise and $R_1 < R < R_2$.*

*Proof.* Without loss of generality assume $a = 0$. As in the proof of Taylor's Theorem, our starting point is Cauchy's Integral Formula.

Fix $z \in A$ and choose $r_1$ and $r_2$ such that $R_1 < r_1 < |z_1| < r_2 < R_2$. Let $\gamma_1 = r_1 e^{it}$ and $\gamma_2 = r_2 e^{it}$, $0 \le t \le 2\pi$, be as shown in Figure 3.41.

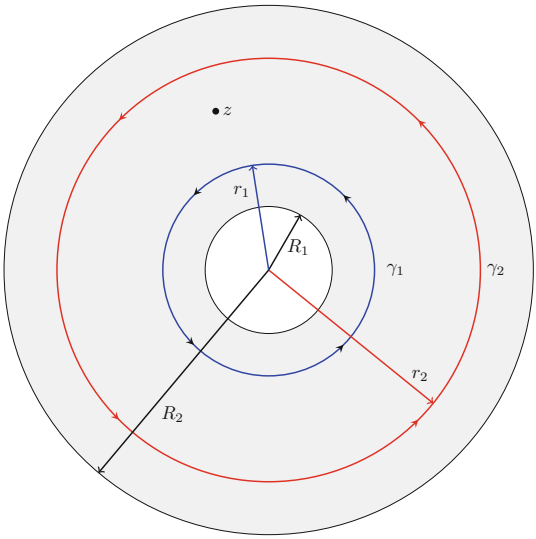



Figure 3.41: Proof of Laurent's theorem

Adding and noting that integrals along the line segments cancel, we have

$$f(z) = \frac{1}{2\pi i}\int_{\gamma_2} \frac{f(w)}{w-z}dw - \frac{1}{2\pi i}\int_{\gamma_1} \frac{f(w)}{w-z}dw. \tag{3.6}$$

Now we can use Taylor's theorem for $w \in \gamma_2$, since $z$ is fixed inside $\gamma_2$; thus $\left|\frac{z}{w}\right| < 1$ and

$$\frac{1}{w-z} = \frac{1}{w\left(1-\frac{z}{w}\right)} = \sum_{n=0}^{\infty} \frac{z^n}{w^{n+1}}.$$

We can use Taylor's theorem again, this time with $w \in \gamma_2$ and $z$ fixed outside $\gamma_1$. Thus $\left|\frac{w}{z}\right| < 1$ and

$$\frac{1}{w-z} = \frac{-1}{-w+z} = -\frac{1}{z}\cdot\frac{1}{1-\left(\frac{w}{z}\right)} = -\sum_{m=0}^{\infty}\frac{w^m}{z^{m+1}}.$$

Returning to equation (3.6), we have

$$f(z) = \frac{1}{2\pi i}\int_{\gamma_2} f(w)\sum_{n=0}^{\infty}\frac{z^n}{w^{n+1}}dw - \frac{1}{2\pi i}\int_{\gamma_1} f(w)\sum_{m=0}^{\infty}\left(-\frac{w^m}{z^{m+1}}\right)dw.$$

We can apply the Weierstrass M-test to obtain uniform convergence in order to interchange the summation and integration operators. This then leads to

$$f(z) = \sum_{n=0}^{\infty} \left( \frac{1}{2\pi i} \int_{\gamma_2} \frac{f(w)}{w^{n+1}} dw \right) z^n + \sum_{m=0}^{\infty} \left( \frac{1}{2\pi i} \int_{\gamma_1} f(w) w^m dw \right) z^{-m-1}.$$

If we then set $m = -(n+1)$ in the second sum, and use the Deformation Theorem to replace $\gamma_1$ and $\gamma_2$ by $\gamma$ in the statement of Laurent's Theorem, we get

$$\begin{aligned} f(z) &= \sum_{n=0}^{\infty} \left( \frac{1}{2\pi i} \int_{\gamma} \frac{f(w)}{w^{n+1}} dw \right) z^n + \sum_{n=-\infty}^{-1} \left( \frac{1}{2\pi i} \int_{\gamma} \frac{f(w)}{w^{n+1}} dw \right) z^n \\ &= \sum_{n=-\infty}^{\infty} \left( \frac{1}{2\pi i} \int_{\gamma} \frac{f(w)}{w^{n+1}} dw \right) z^n, \end{aligned}$$

which gives the desired result. □

**Remark 78.**

a) Note that we cannot set $c_n = \dfrac{f^n(a)}{n!}$ as we did in Taylor's theorem. Indeed $f^{(n)}(a)$ is not even defined since $a \notin A$.

b) (Uniqueness of the coefficients). Let $f \in H(A)$ where $A = \{z : R_1 < |z - a| < R_2\}$ and $R_1 < R_2$ and suppose that $f(z) = \sum_{n=-\infty}^{\infty} b_n (z-a)^n$ for $z \in A$, then one can show that $b_n = c_n$ for all $n$ where $c_n = \dfrac{1}{2\pi i} \displaystyle\int_{\gamma} \frac{f(w)}{(w-a)^{n+1}} dw$ .

**Example 113.** Let $f(z) = \dfrac{1}{z(2-z)}$; suppose we want to find the Laurent expansion of $f$ in

a) $A = \{z : 0 < |z| < 2\}$ and

b) $A = \{z : |z| > 2\}$.

For part a), since $\dfrac{|z|}{2} < 1$, we have

$$f(z) = \frac{1}{z(2-z)} = \frac{1}{2z} \cdot \frac{1}{\left(1 - \frac{z}{2}\right)} = \frac{1}{2z} \sum_{n=0}^{\infty} \left(\frac{z}{2}\right)^n = \sum_{n=0}^{\infty} \frac{z^{n-1}}{2^{n+1}}.$$

For part b), since $\dfrac{2}{|z|} < 1$, we have

$$f(z) = \frac{1}{z(2-z)} = \frac{1}{z^2 \left(\frac{2}{z} - 1\right)} = -\frac{1}{z^2} \sum_{n=0}^{\infty} \left(\frac{2}{z}\right)^n = -\sum_{n=0}^{\infty} \frac{2^n}{z^{n+2}}.$$

## Classification of Singularities

Consider the Laurent expansion of $f$ when $R_1 = 0$. In this case $f$ is holomorphic in the region $\{z : 0 < |z-a| < R_2\}$, which is a deleted neighborhood of $a$. We then say $a$ is an isolated singularity of $f$. Thus

$$f(z) = \cdots + \frac{bn}{(z-a)^n} + \cdots + \frac{b_1}{(z-a)} + a_0 + a_1(z-a) + a_2(z-a)^2 + \cdots$$

is valid for $0 < |z-a| < R$.

**Definition 112.** a) $a$ is called a **removable singularity** if $b_n = 0$ for all $n \geq 0$.

b) $a$ is called a **pole of order** $m$ if $b_m \neq 0$ but $b_n = 0$ for all $n > m$.

c) $a$ is called an **essential singularity** if $b_n \neq 0$ for infinitely many $n$.

**Example 114.**

a) $\dfrac{\sin z}{z}$ has a removable singularity at $z = 0$ because

$$\frac{\sin z}{z} = \frac{z - \dfrac{z^3}{3!} + \dfrac{z^5}{5!} - \cdots}{z} = 1 - \frac{z^2}{3!} + \frac{z^4}{5!} - \cdots$$

for $|z| > 0$.

b) $\dfrac{1-\cos z}{z^3}$ has a simple pole (a pole of order one) at $z = 0$ ($a = 0$) because

$$\frac{1-\cos z}{z^3} = \frac{1 - (1 - \dfrac{z^2}{2!} + \dfrac{z^4}{4!} - \cdots)}{z^3} = \frac{1}{2!} \cdot \frac{1}{z} - \frac{z}{4!} + \frac{z^3}{6!} - \cdots$$

for $|z| > 0$.

c) $e^{1/z^2}$ has an essential singularity at $z = 0$ because

$$\begin{aligned} e^{1/z^2} &= 1 + \frac{1}{z^2} + \frac{\left(\dfrac{1}{z^2}\right)^2}{2!} + \frac{\left(\dfrac{1}{z^2}\right)^3}{3!} + \cdots + \frac{\left(\dfrac{1}{z^2}\right)^n}{n!} + \cdots \\ &= 1 + \frac{1}{z^2} + \frac{1}{2!} \cdot \frac{1}{z^4} + \frac{1}{3!} \cdot \frac{1}{z^6} + \cdots + \frac{1}{n!} \cdot \frac{1}{z^{2n}} + \cdots . \end{aligned}$$

Now we give a theorem of identification, for more details consult [54].

**Theorem 146.** *Suppose $z = a$ is an isolated singularity of the function $f(z)$. Then*

*a) $a$ is a removable singularity* $\iff \lim_{z\to a}(z-a)f(z) = 0$.

*b) i) $a$ is a pole of order $m$* $\iff \lim_{z\to a}(z-a)^m f(z) = l \neq 0$.

*ii) $a$ is a pole of order 1 (or $a$ is a simple pole)* $\iff \lim_{z\to a} f(z) = \infty$.

*c) $a$ is an essential singularity* $\iff \lim_{z\to a} f(z)$ *does not exist (i.e., $f(z)$ tends to no limit, finite or infinite).*

**Example 115.**

a) $\dfrac{\sin z}{z}$ has a removable singularity at $z = 0$ because

$$\lim_{z\to 0}(z-0)\frac{\sin z}{z} = \sin(0) = 0 \text{ (or } \lim_{z\to 0}\frac{\sin z}{z} = 1).$$

b) $\dfrac{1-\cos z}{z^3}$ has a pole of order 1 at $z = 0$ since

$$\lim_{z\to 0}\frac{1-\cos z}{z^3} = \lim_{z\to 0}\frac{-\sin z}{3z^2} = \lim_{z\to 0}\frac{\cos z}{6z} \to \infty.$$

c) $f(z) = \dfrac{5z+3}{(1-z)^3\sin^2 z}$ where $0 < |z| < 1$ has a pole of order 3 at $z = 1$ and a pole of order 2 at $z = 0$ because

$$\lim_{z\to 1}(z-1)^3\frac{5z+3}{(1-z)^3\sin^2 z} = \frac{-8}{\sin^2(1)} \neq 0$$

and

$$\lim_{z\to 0}(z-0)^2 f(z) = \lim_{z\to 0} z^2 \cdot \frac{5z+3}{(1-z)^3\sin^2 z} = 3.$$

d) $e^{1/z^2}$ has an essential singularity at $z = 0$ because $\lim_{z\to 0} e^{1/z^2}$ does not exist. If $z = x$, then

$$\lim_{z\to 0} e^{1/z^2} = \lim_{x\to 0} e^{1/x^2} \to \infty$$

while if $z = iy$, then

$$\lim_{z\to 0} e^{1/z^2} = \lim_{y\to 0} e^{1/(iy)^2} = \lim_{y\to 0}\frac{1}{e^{1/y^2}} = 0.$$

## Calculation of Residues

**Notation.** Suppose $f$ is holomorphic in the punctured disc $D'(a;r) = \{z \in \mathbb{C} : 0 < |z-a| < r\}$ and that $f$ has a pole at $a$. The **residue** of $f$ at $a$ is denoted by $\text{Res}(f(z), a)$, and defined to be the coefficient $b_1$ of $(z-a)^{-1}$,

$$\text{Res}(f(z), a) = b_1.$$

**Theorem 147.**

*a) If $f(z)$ has a removable singularity at $a$, then*

$$Res(f(z), a) = 0.$$

*b) If $a$ is a simple pole of $f(z)$, then*

$$Res(f(z), a) = \lim_{z \to a}(z-a)f(z)$$

*c) If $a$ is a pole of order $k$ ($k \geq 2$) for $f(z)$, then*

$$Res(f(z), a) = \frac{1}{(k-1)!} \lim_{z \to a} \left[ \frac{d^{k-1}}{dz^{k-1}} [(z-a)^k f(z)] \right].$$

*Proof.*

a) If $a$ is a removable singularity of $f(z)$, then $f(z) = \sum_{n=0}^{\infty} a_n (z-a)^n$. Thus there is no coefficient $b_1$, so $\text{Res}(f(z), a) = 0$.

b) If $a$ is a simple pole, then $f(z) = \dfrac{b_1}{z-a} + \sum_{n=0}^{\infty} a_n (z-a)^n$. Therefore

$$(z-a)f(z) = b_1 + \sum_{n=0}^{\infty} a_n (z-a)^{n+1}$$

and

$$\lim_{z \to a}(z-a)f(z) = b_1.$$

c) If $a$ is a pole of order $k$, for $f(z)$, then

$$f(z) = \frac{b_k}{(z-a)^k} + \frac{b_{k-1}}{(z-a)^{k-1}} + \cdots + \frac{b_1}{(z-a)} + \sum_{n=0}^{\infty} a_n (z-a)^n \quad \text{with } b_k \neq 0.$$

The goal here is to obtain the coefficient $b_1$. To show this formula does in fact yield $b_1$, observe that

$$(z-a)^k f(z) = b_k + b_{k-1}(z-a) + b_{k-2}(z-a)^2 + \cdots + b_1 (z-a)^{k-1} + \sum_{n=0}^{\infty} a_n (z-a)^{n+k}.$$

Now if we differentiate this $k-1$ times, only the terms with an order $k-1$ or higher survive. That is, we get

$$\frac{d^{k-1}}{dz^{k-1}}\left[b_k + b_{k-1}(z-a) + b_{k-2}(z-a)^2 + \cdots + b_1(z-a)^{k-1} + \sum_{n=0}^{\infty} a_n(z-a)^{n+k}\right]$$
$$= 0 + 0 + 0 + \cdots + b_1(k-1)! + \sum_{n=0}^{\infty}(n+k)(n+k-1)\cdots(n+2)a_n(z-a)^{n+1}.$$

Finally, as we take $z \to a$, we see that

$$\lim_{z\to a}\left[b_1(k-1)! + \sum_{n=0}^{\infty}(n+k)(n+k-1)\cdots(n+2)a_n(z-a)^{n+1}\right] = b_1(k-1)!.$$

That is, the power series vanishes. If we finally divide by $(k-1)!$, we'll obtain $b_1$. Overall, the way we obtain $b_1$ is by multiplying through by $(z-a)^k$, differentiating the result $(k-1)$ times and taking the limit as $z \to a$ to obtain

$$b_1 = \text{Res}(f(z), a) = \frac{1}{(k-1)!}\lim_{z\to a}\left[\frac{d^{k-1}}{dz^{k-1}}[(z-a)^k f(z)]\right]. \qquad \square$$

**Remark 79.** Let $f(z) = \dfrac{g(z)}{h(z)}$ where $g$ and $h$ are holomorphic at $z = a$. Furthermore, suppose $g(a) \neq 0$ and $h(a) = 0$, but $h'(a) \neq 0$ ( in such case we say $h$ has a zero of order 1 at $a$). Then $f$ has a simple pole at $z = a$, and

$$\text{Res}(f(z), a) = \frac{g(a)}{h'(a)}.$$

This easy way of finding residues for such functions $f(z) = \dfrac{g(z)}{h(z)}$ follows from the fact that

$$\text{Res}(f(z), a) = \lim_{z\to a}(z-a)\frac{g(z)}{h(z)} = \lim_{z\to a}\frac{g(z)}{\frac{h(z)}{z-a}} = \frac{g(a)}{h'(a)}.$$

**Example 116.**

a) $f(z) = \dfrac{1}{1+z^2} = \dfrac{1}{(z+i)(z-i)}$ has simple poles at $z = i$ and $z = -i$. Find the residues there.

*Solution.* To find the residues, we calculate

$$\text{Res}(f(z), i) = \lim_{z\to i}(z-i)f(z) = \lim_{z\to i}(z-i)\frac{1}{(z+i)(z-i)} = \frac{1}{2i}.$$

and

$$\text{Res}(f(z), -i) = \lim_{z\to -i}(z+i)f(z) = \lim_{z\to -i}(z+i)\frac{1}{(z+i)(z-i)} = -\frac{1}{2i}.$$

The residues at $z = i$ and $z = -i$ are $\dfrac{1}{2i}$ and $-\dfrac{1}{2i}$, respectively.

b) $f(z) = \dfrac{e^z}{z^3}$ has a pole of order 3 at $z = 0$ since $\lim_{z\to 0} z^3 \cdot \dfrac{e^z}{z^3} = 1 \neq 0$. Find the residues at the pole.

*Solution.* To find the residue, we calculate

$$\operatorname{Res}(f(z), 0) = \frac{1}{2!} \lim_{z\to 0} \left[ \frac{d^2}{dz^2}(z-0)^3 f(z) \right] = \frac{1}{2!} \lim_{z\to 0} \left[ \frac{d^2}{dz^2} z^3 \cdot \frac{e^z}{z^3} \right] = \frac{1}{2}.$$

Thus the residue of the pole at $z = 0$ is $\frac{1}{2}$.

c) Calculate the residues of the function $f(z) = \dfrac{\sin z}{(z-i)(z+2)^2}$ at its singularities.

*Solution.* Observe that $z = i$ and $z = -2$ are singularities for $f(z)$, and that $z = i$ is a simple pole while $z = -2$ is a pole of order 2. Therefore,

$$\operatorname{Res}(f(z), i) = \lim_{z\to i}(z-i)\frac{\sin z}{(z-i)(z+2)^2} = \frac{\sin i}{(i+2)^2}$$

and

$$\begin{aligned} \operatorname{Res}(f(z), -2) &= \lim_{z\to -2} \frac{d}{dz}\left[ (z+2)^2 \frac{\sin z}{(z-i)(z+2)^2} \right] = \lim_{z\to -2} \frac{d}{dz}\left[ \frac{\sin z}{(z-i)} \right] \\ &= \lim_{z\to -2} \frac{(z-i)\cos z - \sin z}{(z-i)^2} = \frac{(-2-i)\cos(2) + \sin(2)}{3+4i}. \end{aligned}$$

d) The function $f(z) = \dfrac{\sin z - z}{z^4}$ has a pole at $z = 0$, but the order of this pole is not immediately evident. However, if we write $\sin z - z = -\dfrac{z^3}{3!} + \dfrac{z^5}{5!} - \cdots$ then

$$f(z) = \frac{-\frac{z^3}{3!} + \frac{z^5}{5!} - \cdots}{z^4} = -\frac{1}{3!z} + \frac{z}{5!} - \cdots.$$

Thus we have that $\operatorname{Res}(f(z), 0) = -\dfrac{1}{3!} = -\dfrac{1}{6}$.

## Cauchy's Residue Theorem

**Theorem 148.** *Let $f$ be holomorphic inside and on a positively oriented contour except for a finite number of poles $a_1, a_2, \ldots, a_n$ inside $\gamma$ (Figure 3.42). Then*

$$\int_\gamma f(z)dz = 2\pi i \sum_{k=1}^{n} Res(f(z), a_k).$$

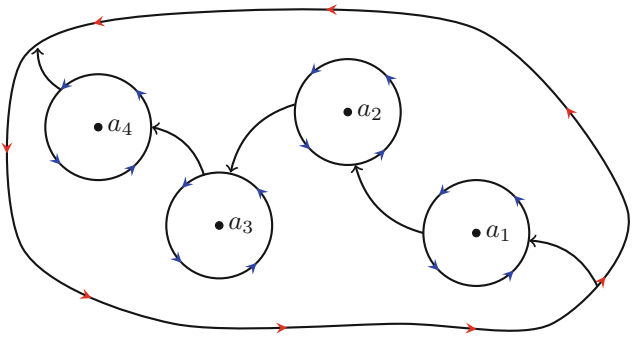



Figure 3.42: $f$ is not holomorphic for finite number of poles $a_1, a_2, a_3, a_4$ in $\gamma$

*Proof.* Consider the Laurent expansion at $a_k$:

$$f(z) = \underbrace{\frac{b_m}{(z-a_k)^m} + \cdots + \frac{b_1}{(z-a_k)^m}}_{\text{Call this part } f_k(z)} + \sum_{n=0}^{\infty} a_n (z-a_k)^n.$$

Define a function $g := f(z) - \sum_{k=1}^{N} f_k(z)$. Then $g(z)$ has only removable singularities at $a_1, a_2, \ldots, a_N$. Remove them and use Cauchy's theorem to deduce that $\int_\gamma g(z)dz = 0$. But

$$\int_\gamma g(z)dz = \int_\gamma f(z)dz - \sum_{k=1}^{N} \int_\gamma f_k(z)dz = 0.$$

Hence,

$$\int_\gamma f(z)dz = \sum_{k=1}^{N} \int_\gamma f_k(z)\, dz.$$

Note that the integral around a pole can be found easily using the fundamental integral and we have

$$\int_\gamma f_k(z)dz = 2\pi i \mathrm{Res}(f(z), a_k)$$

(see Exercise 17 below). Therefore,

$$\int_\gamma f(z)dz = 2\pi i \sum_{k=1}^{N} \mathrm{Res}(f(z), a_k).$$

□

**Applications of Cauchy's Residue Theorem**

**Example 1.** Use Cauchy's Residue Theorem to evaluate

$$\int_\gamma \frac{1-2z}{z(1-z)(z-3)}dz$$

where $\gamma : |z| = 2$.

*Solution.* The integrand $f(z) = \dfrac{1-2z}{z(1-z)(z-3)}$ has simple poles at $z = 0, z = 1$ and $z = 3$. However, only the first two of these points lie inside $\gamma$. Thus by the Residue Theorem,

$$\int_\gamma f(z)dz = 2\pi i[\text{Re}(f(z), 0) + \text{Re}(f(z), 1)].$$

Since

$$\text{Re}(f(z), 0) = \lim_{z\to 0} zf(z) = \lim_{z\to 0} \frac{1-2z}{(1-z)(z-3)} = -\frac{1}{3} \quad \text{and}$$

$$\text{Re}(f(z), 1) = \lim_{z\to 1}(z-1)f(z) = \lim_{z\to 1} \frac{-1+2z}{z(z-3)} = -\frac{1}{2},$$

we obtain that

$$\int_\gamma f(z)dz = 2\pi i\left(\frac{-1}{2} + \frac{-1}{3}\right) = -\frac{5\pi i}{3}.$$

**Example 2.** Use Cauchy's Residue Theorem to evaluate $\displaystyle\int_\gamma \frac{dz}{1+z^4}$ where $\gamma : [-R, R] \cup \Gamma$ as shown in Figure 3.43.

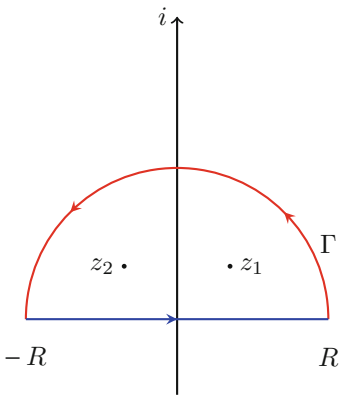



Figure 3.43: Only $z_1$ and $z_2$ are inside the given contour $\Gamma$

*Solution.* Observe that $1 + z^4 = 0 \implies z^4 = -1$. Thus the roots are

$$z_1 = e^{i\pi/4}, \quad z_2 = e^{i3\pi/4}, \quad z_3 = e^{i5\pi/4}, \quad z_4 = e^{i7\pi/4}.$$

Only $z_1$ and $z_2$ are inside the given contour. Therefore,

$$\int_\gamma \frac{dz}{1+z^4} = 2\pi i \left[\text{Res}(f(z), z_1) + \text{Res}(f(z), z_2)\right].$$

Now

$$\text{Res}(f(z), z_1) = \lim_{z \to e^{i\pi/4}} \frac{z - e^{\pi/4}}{1+z^4} = \lim_{z \to e^{i\pi/4}} \frac{1}{4z^3} = \frac{1}{4} e^{-3\pi i/4} \quad \text{and}$$

$$\text{Res}(f(z), z_2) = \lim_{z \to e^{3\pi i/4}} \frac{z - e^{3\pi i/4}}{1+z^4} = \lim_{z \to e^{3\pi i/4}} \frac{1}{4z^3} = \frac{1}{4} e^{-\pi i/4}.$$

Therefore

$$\int_\gamma \frac{dz}{1+z^4} = 2\pi i \left[\frac{1}{4} e^{-3\pi i/4} + \frac{1}{4} e^{-\pi i/4}\right] = \frac{\pi}{\sqrt{2}}.$$

**Example 3.** Use Cauchy's Residue Theorem to evaluate the following real trigonometric integral

$$\int_0^{2\pi} \frac{\sin^2(\theta)}{5 + 4\cos(\theta)} d\theta.$$

*Solution:* First we establish the identification

$$z = e^{i\theta} \quad 0 \le \theta \le 2\pi.$$

With this choice of identification, we have that $\dfrac{1}{z} = e^{-i\theta}$. Since

$$\cos\theta = \frac{e^{i\theta} + e^{-i\theta}}{2} = \frac{z + \dfrac{1}{z}}{2} \quad \text{and} \quad \sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i} = \frac{z - \dfrac{1}{z}}{2i},$$

we have that

$$\cos\theta = \frac{1}{2}\left(z + \frac{1}{z}\right) \quad \text{and} \quad \sin\theta = \frac{1}{2i}\left(z - \frac{1}{z}\right).$$

Furthermore, when integrating along $\gamma$,

$$z = e^{i\theta} \implies dz = iz d\theta \implies \frac{dz}{iz} = d\theta.$$

Secondly, observe that $5 + 4\cos\theta$ is never zero. Thus the integrand is finite over $[0, 2\pi]$. Performing the above substitutions leads to

$$\int_0^{2\pi} \frac{\sin^2\theta}{5 + 4\cos\theta} d\theta = \int_\gamma \frac{\left[\dfrac{1}{2i}\left(z - \dfrac{1}{z}\right)\right]^2}{5 + 4 \cdot \dfrac{1}{2}\left(z + \dfrac{1}{z}\right)} \frac{dz}{iz},$$

which can be reduced to

$$-\frac{1}{4i}\int_\gamma \frac{(z^2-1)^2}{z^2(2z^2+5z+2)}dz = -\frac{1}{4i}\int_\gamma \frac{(z^2-1)^2}{2z^2\left(z+\frac{1}{2}\right)(z+2)}dz.$$

Let $g(z) = \dfrac{(z^2-1)^2}{2z^2\left(z+\frac{1}{2}\right)(z+2)}$. Observe that $g(z)$ has simple poles at $z=-\frac{1}{2}$, $z=-2$ and a pole of order 2 at $z=0$. However, only $-\dfrac{1}{2}$ and 0 lie inside the unit circle $\gamma$, so that

$$\int_0^{2\pi} \frac{\sin^2\theta}{5+4\cos\theta}d\theta = -\frac{1}{4i}\cdot 2\pi i\left[\text{Res}(g(z),-\frac{1}{2}) + \text{Res}(g(z),0)\right].$$

Now we calculate the residues as follows:

$$\text{Res}\left(g(z),-\frac{1}{2}\right) = \lim_{z\to\frac{1}{2}}[(z+\frac{1}{2})\cdot g(z)] = \lim_{z\to-\frac{1}{2}}\frac{(z^2-1)^2}{2z^2(z+2)} = \frac{3}{4}$$

and

$$\text{Res}\,(g(z),0) = \lim_{z\to0}\frac{1}{1!}\frac{d}{dz}[z^2g(z)] = \lim_{z\to0}\frac{d}{dz}\left[\frac{(z-1)^2}{2z^2+5z+2}\right]$$
$$= \lim_{z\to0}\frac{(2z^2+5z+2)\cdot 2(z^2-1)2z-(z-1)^2(4z+5)}{(2z^2+5z+2)^2} = -\frac{5}{4}.$$

Hence,

$$\int_0^{2\pi} \frac{\sin^2\theta}{5+4\cos\theta}d\theta = -\frac{1}{4i}\cdot 2\pi i\left[\frac{3}{4}-\frac{5}{4}\right] = \frac{\pi}{4}.$$

## Exercises

1. Find the Laurent expansion about $a=0$ which represents the following functions in the indicated regions.

a) $\dfrac{z^2-1}{(z+2)(z+3)}$ for $|z|>3$.

Hint: $\dfrac{z^2-1}{(z+2)(z+3)} = 1-\dfrac{5z+7}{(z+2)(z+3)} = 1+\dfrac{A}{z+2}+\dfrac{B}{z+3}$

with $A=3$, $B=-8$.

b) $\dfrac{24}{z^2(z-1)(z+2)}$ for $0 < |z| < 1$.

Hint : $\dfrac{24}{z^2(z-1)(z+2)} = \dfrac{A}{z} + \dfrac{B}{z^2} + \dfrac{C}{z-1} + \dfrac{D}{z+2}$

where $A = -6$, $B = -12$, $C = 8$, $D = -2$.

2. Find the Laurent series about $a = 1$ and then $a = 0$ for $f(z) = \dfrac{1}{z^2 - z^3}$.

3. Find the principal part and residue of $f(z) = \dfrac{z^3 + z^2}{(z-1)^2}$ at $a = 1$.
Hint: Expand $z^3+z^2$ in powers of $z-1$: $z^3+z^2 = 2+5(z-1)+4(z-1)^2+(z-1)^3$.

4.

a) For $f(z) = e^{\frac{1}{z^2}}$, for $|z| > 0$, show that $a = 0$ is an isolated essential singularity.

b) For $f(z) = \dfrac{\sin z}{z^4}$, $|z| > 0$, show that $a = 0$ is a pole of order 3.

5. Locate and classify all singularities of the following functions:

a) $f(z) = \dfrac{2}{(z-3)^2} + \dfrac{1}{z-3} + e^z$.

b) $f(z) = \sin z + \sin \dfrac{1}{z}$.

c) $f(z) = \dfrac{\cos z}{z - \dfrac{\pi}{2}}$.

6. Show that the function $f(z) = \dfrac{z^3 - 8}{(z-2)^2(z+i)^4}$ has a simple pole at 2 and a pole of order 4 at $-i$.

7. Let $f$ be holomorphic on $\mathbb{C} \setminus \{0\}$. Show that the Laurent expansions for $f$ valid in the regions $\{z : |z| > 0\}$ and $\{z : |z| > 1\}$ are the same.

8. Find the Laurent series for $f(z) = \dfrac{8z+1}{z(1-z)}$ valid for $0 < |z| < 1$.

9. Let $f$ and $g$ be continuous on $\overline{A}$ and holomorphic on $A$, where $A$ is open, i.e., a connected and bounded region. If $f = g$ on $\partial A$, show that $f = g$ on all of $\overline{A}$.

10. If $f$ is entire and bounded on the real axis, then $f$ is constant. Prove or give a counterexample.

11. Find the Laurent series for

$$f(z) = \frac{4}{(1-z)(z+3)}$$

in the annulus $\{z : \ 1 < |z| < 3\}$.

12. Find the residue of the following functions at $z = 0$

a) $f(z) = \dfrac{z^2+1}{z}$,

b) $f(z) = \dfrac{\sin z}{z^4}$.

13. Evaluate:

a) $\displaystyle\int_\gamma \operatorname{Re} z dz$ where $\gamma(t) = |z| = 1$,

b) $\displaystyle\int_0^{3+i} \sin z dz$,

c) $\displaystyle\int_\gamma \frac{z+4}{z^4+2iz^3}\, dz$ where $\gamma$ is the circle $|z| = 1$.

14. Use the residue theorem to evaluate the following:

a) $\displaystyle\int_\gamma e^{4/z-2}\ dz$ where $\gamma$ is $|z-1| = 3$,

b) $\displaystyle\int_\gamma \frac{e^z}{z^3+2z^2}\, dz$ where $\gamma$ is $|z| = 3$,

c) $\displaystyle\int_0^{2\pi} \frac{1}{10-6\cos\theta}\ d\theta$.

15. Let $f, g$ be holomorphic functions on $D(a; r)$. Assume that $f$ has a zero of order $m$ while $g$ has a zero of order $m+1$ at $a$. Show that

$$\operatorname{Res}\left(\frac{f(z)}{g(z)}; a\right) = (m+1)\frac{f^m(a)}{g^{(m+1)}(a)}.$$

16. Evaluate $\displaystyle\int_\gamma \tan z\, dz$ where $\gamma = |z| = 2$.

17. Let $f$ be holomorphic inside and on a positively oriented contour $\gamma$ except at the point $a$ inside $\gamma$, where it has a pole of order $m$. Let $\displaystyle\sum_{n=-m}^{\infty} c_n(z-a)^n$ be the Laurent expansion of $f$ about $a$. Show that $\displaystyle\int_\gamma f(z)dz = 2\pi i c_{-1}$.

18. Find the Laurent series that converges in the annulus $1 < |z| < 2$ to a branch of the function $\log\left(\frac{z(2-z)}{1-z}\right)$.

19. Show that $\displaystyle\int_{-\infty}^{\infty} \frac{e^{ax}}{1+e^x}\,dx = \frac{\pi}{\sin ax}$ for $0 < a < 1$.

Hint: Consider $f(z) = \dfrac{e^{az}}{1+e^z}$ use residue theorem over the contour

$$\gamma = [-R,\, R] \cup \gamma_1 \cup \gamma_2 \cup \gamma_3$$

as shown in the following Figure 3.44.

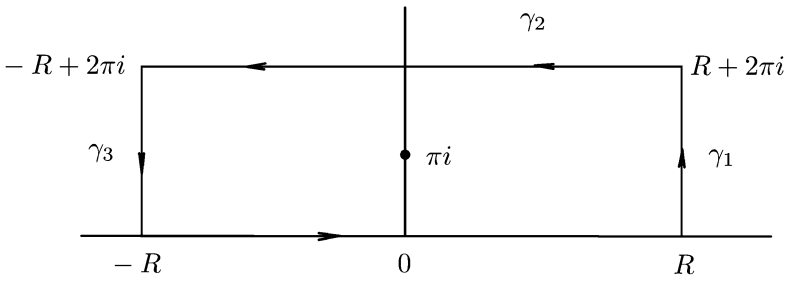



Figure 3.44: Rectangular contour

## 3.10 The Bieberbach Conjecture

One of the most celebrated conjectures in classical analysis which stood as a challenge to mathematicians for nearly 70 years is called the Bieberbach conjecture. This conjecture appeared in a footnote to a paper [11] of a German mathematician, Ludwig Bieberbach, in 1916 and was solved by Louis de Branges of Purdue University in 1984 [19]. The Bieberbach conjecture is appealing partly because it is simple to pose, and it states that under reasonable restrictions the coefficients of a power series are not too large. The Bieberbach conjecture concerns functions which are both holomorphic and also univalent. A holomorphic function is *univalent* if it is one-to-one ($f(z_1) \neq f(z_2)$ unless $z_1 = z_2$). Univalent functions have many interesting properties [20], thus we may wonder if we can say anything about the coefficients in their Taylor expansions. It turns out that if we also assume $f(0) = 0$ and $f'(0) = 1$ then the Taylor series for $f$ takes the form

$$f(z) = z + a_2 z^2 + a_3 z^3 + \cdots$$

with complex coefficients $a_2, a_3, \ldots$. We use the letter $S$ (for Schlicht) for the class of univalent and holomorphic functions $f : \mathbb{D} \to \mathbb{C}$ which satisfy the normalization conditions $f(0) = 0$ and $f'(0) = 1$.

**Bieberbach Conjecture** : For each $f \in S$, $|a_n| \leq n$ for $n = 2, 3, \ldots$ The inequality is strict for every $n$ unless $f$ is a rotation of the Koebe function $k(z)$ where

$$k(z) = \sum_{n=1}^{\infty} nz^n = z + 2z^2 + 3z^3 + \cdots .$$

The principal result of Bieberbach's original paper was the second coefficient theorem, $|a_2| \leq 2$, and that equality holds for the Koebe function. Before de Branges' general proof of $|a_n| \leq n$, this conjecture was known to be true only for $n \leq 6$. There are several books and papers written on this conjecture; we refer the reader to [24, 55] for more detailed information.
The Koebe function $k(z)$ mentioned in this conjecture is

$$k(z) = \frac{z}{(1-z)^2} = z\,\frac{d}{dz}\left[\frac{1}{(1-z)}\right] = z + 2z^2 + 3z^3 + \cdots$$

which converges for every $z$ in the disc $|z| < 1$. To see why $k(z)$ is univalent on the disc and to find its image, consider $k(z)$ in the following form:

$$k(z) = \frac{1}{4}\left[\left(\frac{1+z}{1-z}\right)^2 - 1\right].$$

We see that $k(z)$ is composition of the following mappings

$$p = \frac{1+z}{1-z}, \qquad q = p^2, \qquad w = \frac{1}{4}(q-1)$$

First $p$ is a linear fractional transformation that maps the unit disc univalently onto the right half of the $p$-plane. The mapping $q = p^2$ is one-to-one when restricted to the right half-plane; its image is the entire $q$ plane minus the non-negative real axis. Finally, the last mapping $w$ is a simple translation followed by a dilation with a factor of $\frac{1}{4}$ as shown in Figure 3.45.

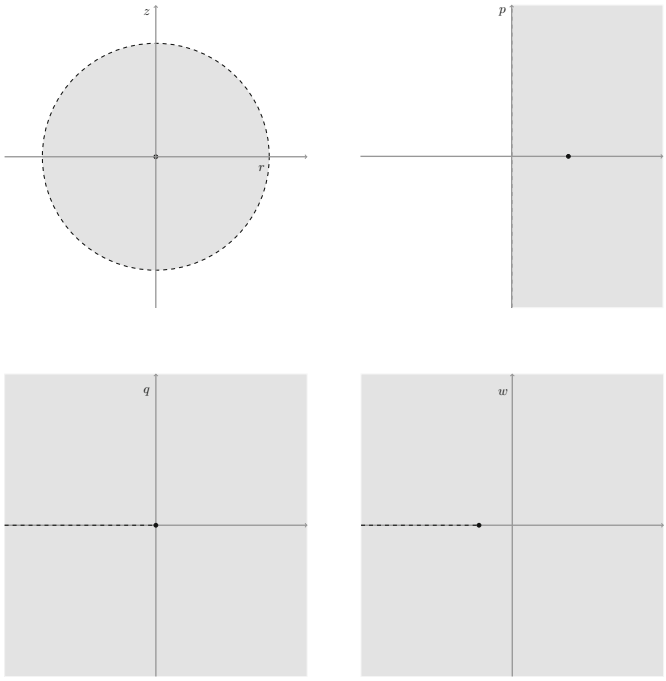



Figure 3.45: Koebe function

Given a function $f$ satisfying $f(0) = 0,\ f'(0) = 1$ and a real number $\alpha$, then

$$g(z) = e^{-i\alpha} f(e^{i\alpha} z)$$

is a counterclockwise rotation of $f$ about $z = 0$ through $\alpha$ radians. Expressing it as power series, we get that if

$$f(z) = z + \sum_{n=2}^{\infty} a_n z^n \quad \text{then} \quad g(z) = z + \sum_{n=2}^{\infty} b_n z^n,$$

where $b_n = a_n e^{i(n-1)\alpha}$. Therefore

$$|b_n| = |a_n||e^{i(n-1)\alpha}| = |a_n|.$$

These functions

$$k_\lambda(z) = \frac{z}{(1-\lambda z)^2} = \sum_{n=1}^{\infty} n\lambda^{n-1} z^n \quad (\lambda \quad \text{a constant and} \quad |\lambda| = 1)$$

are, in fact, the only functions for which $|a_n| = n$ for some (and hence for all) $n$, up to rotations. In the following example $f(z)$ is a polynomial of degree $n$ and we try to find a certain bound on $|a_n|$.

**Example 117.** Given

$$f(z) = z + a_2 z^2 + a_3 z^3 + \cdots + a_n z^n$$

a polynomial of degree $n$, if $f$ is univalent in $D(0;1)$, then $|a_n| \leq \frac{1}{n}$. To see this, consider

$$f'(z) = 1 + 2a_2 z + \cdots + n a_n z^{n-1} = n a_n (\frac{1}{n a_n} + \cdots + z^{n-1}).$$

Now use the Fundamental Theorem of Algebra to factor the above polynomial of degree $n-1$ as

$$f'(z) = n a_n (z - c_1)(z - c_2) \cdots (z - c_{n-1})$$

where $c_1, c_2, \ldots, c_{n-1}$ are complex roots. However, $f$ is univalent on $D(0;1)$, so $f'(z)$ has no roots in $D(0;1)$ and thus $|c_k| > 1$ for all $k$. Since $f'(0) = 1$,

$$1 = |f'(0)| = |n a_n||c_1||c_2| \cdots |c_{n-1}| \geq |n a_n|$$

as claimed.

The most important property of univalent functions is the famous Riemann mapping theorem which we state below again. See also Theorem 127.

**Theorem 149** (Riemann Mapping Theorem)**.** *Let $G$ be a simply connected region ("without holes") with $G \neq \mathbb{C}$. Then there exists a one-to-one conformal mapping $f : G \to D(0,1)$ with $f^{-1} : D(0,1) \to G$ is also conformal.*

The Riemann mapping theorem is an existence result; it does tell you how you can find such a mapping. It is not clear at all that it is possible to construct a conformal map from a region with complicated boundary onto a nice region such as $D(0;1)$ or vice versa. The strength of the Riemann mapping theorem is the reason why the Bieberbach conjecture stayed a main problem in geometric function theory till de Branges' proof.

## Bieberbach's Area Theorem

The Bieberbach conjecture came to be postulated after the proof of the following area theorem. The basic way to obtain Bieberbach like inequalities is to relate the power series coefficients to the area of some region in the plane. Let $\mathcal{U}$ denote the class of functions $g$ which are holomorphic and univalent in $\Sigma = \{z : |z| > 1\}$ and have Laurent expansion

$$g(z) = z + b_0 + \frac{b_1}{z} + \cdots = z + \sum_{n=0}^{\infty} \frac{b_n}{z^n}.$$

**Theorem 150.** *If $g(z)$ is in $\mathcal{U}$, then* $\sum_{n=1}^{\infty} n|b_n|^2 \leq 1.$

*Proof.* Let $R = \mathbb{C} \setminus g(\Sigma)$ be the complement of the image set $g(\Sigma)$. Our goal is to calculate the area of $R$ in terms of the $b_n$. More precisely, we set $R(r) = \mathbb{C} \setminus \{g(z) : |z| > r\}$ and find the area bounded by the simple closed curve $\gamma(r)$ (here by $\gamma(r)$ we denote the boundary of $R(r)$.) Then we define (Figure 3.46):

$$\operatorname{area} R = \lim_{r \to 1^+} \operatorname{area} R(r).$$

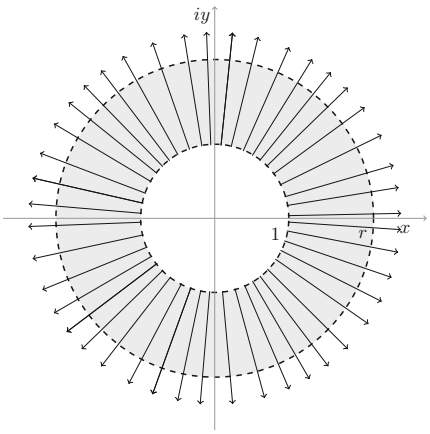



Figure 3.46: Area of R

Let $g(z) = u + iv = w$; clearly,

$$\operatorname{area} R(r) = \iint_{R(r)} du\, dv = \frac{1}{2i} \iint_{R(r)} d\overline{w}\, dw,$$

since $du\, dv = \dfrac{1}{2i} d\overline{w}\, dw$. Now, we use Green's theorem to write

$$\operatorname{area} R(r) = \frac{1}{2i} \int_{\gamma(r)} \overline{w}\, dw = \frac{1}{2i} \int_{|z|=r} \overline{g(z)} g'(z)\, dz.$$

Next we parametrize $z = re^{it}$, $dz = ire^{it} dt$ to express the last integral as:

$$\operatorname{area} R(r) = \frac{1}{2} \int_0^{2\pi} re^{it} g' r(e^{it}) \overline{g(re^{it})} dt.$$

Using the power series for $g$ and $g'$,

$$\frac{1}{\pi} \operatorname{area} R(r) = \frac{1}{2\pi} \int_0^{2\pi} \left( re^{it} + \sum_{n=1}^{\infty} n b_n r^{-n} e^{-int} \right) \left( re^{-it} + \sum_{n=0}^{\infty} \overline{b_n} r^{-n} e^{int} \right) dt.$$

The last expression can be simplified using the orthogonality of distinct powers of $e^{it}$ to

$$\frac{1}{\pi}\text{area}\, R(r) = r^2 - \sum_{n=1}^{\infty} r^{-2n}\, n\, |b_n|^2.$$

Noting that $\text{area}\, R(r) \geq 0$, and thus the partial sums $\sum_{n=1}^{m} r^{-2n}\, n\, |b_n|^2 \leq r^2$ for every $m > 0$ and letting $r \to 1^+$ yield the claim that $\sum_{n=1}^{m} n|b_n|^2 \leq 1$, for $m = 1, 2, \ldots$. □

This theorem gives us an immediate corollary, useful in proving the Bieberbach theorem:

**Corollary 27.** *If $g(z)$ is in $\mathcal{U}$, then $|b_1| \leq 1$, with equality if and only if $g$ has the form*

$$g(z) = z + b_0 + \frac{b_1}{z} \qquad |b_1| = 1.$$

## The Koebe $\frac{1}{4}$-Theorem

Another well-known result, coming from the case $n = 2$ in the Bieberbach conjecture is known as the Koebe $\frac{1}{4}$ -theorem. This theorem states that a univalent function $f$ on $D(0;1)$ with the conditions $f(0) = 0$ and $f'(0) = 1$ is one such that

$$D(0, \frac{1}{4}) \subseteq f\left(D(0;1)\right).$$

Equality occurs for those class of functions that give equality in the Bieberbach conjecture. Namely, equality can hold if and only if $f$ is a rotation of the Koebe function. In other words, the range of each function $f \in S$ contains some disc centered at the origin. Already in 1907, Koebe conjectured that $r \leq \frac{1}{4}$, with the maximum attained by the Koebe function. Later in 1916, Bieberbach proved the theorem, showing that this constant cannot be improved.

**Remark 80.** Progress to prove the Bieberbach conjecture occurred in several directions; some can be grouped as follows:

a) $|a_n| \leq n$ for a particular $n$,

b) $|a_n| \leq n$ for subclasses of $S$,

c) $|a_n| \leq C$ for sufficiently large $C$

One of the results of the first type is the third coefficient theorem, $|a_3| \leq 3$, due to K. Loewner. His proof was completely different than Bieberbach's second coefficient theorem and he used a partial differential equations method which was useful to de Branges's proof of the general conjecture. The Bieberbach conjecture also had been proved for certain subclasses of S. For example, the

subclass of starlike functions, and the subclass of functions with real coefficients [20]. As an evidence of the difficulty of solving Bieberbach conjecture one can use the basic but powerful tools of analysis (Cauchy's inequality) to show for $f \in S$,

$$|a_n| \leq \frac{e^2}{4} n^2 \quad n = 2, 3, \ldots$$

See [55] for details. Note that the above inequality does not even get the order of growth of the coefficients right. As an example for the third group of results we present the following work of Littlewood.

## Littlewood Theorem

In 1925 Littlewood proved a different, looser bound. While Bieberbach conjectured that each coefficient is bound above by its index ($|a_n| \leq n$), Littlewood proved that for all $n$, $|a_n| \leq e\, n$.

**Theorem 151.** *Let $f(z) \in S$. Then $|a_n| < e\, n$ for $n \geq 2$.*

*Proof.* The proof depends on Littlewood's integral inequality [20] which is: for $f \in S$

$$\frac{1}{2\pi} \int_0^{2\pi} |f(re^{i\theta}| d\theta \leq \frac{r}{1-r} \quad 0 \leq r < 1.$$

Using Cauchy's integral formula,

$$|a_n| = \left| \frac{f^{(n)}(0)}{n!} \right| = \left| \frac{1}{2\pi} \int_{|z|=r} \frac{f(z)}{z^{n+1}}\, dz \right|, \qquad r < 1.$$

Changing $z = re^{i\theta}$ and using Littlewood's integral inequality we obtain

$$|a_n| = \left| \frac{1}{2\pi} \int_0^{2\pi} \frac{f(re^{i\theta})}{r^n \cdot e^{in\theta}}\, dz \right| \leq \frac{1}{2\pi r^n} \int_0^{2\pi} \left| f(re^{i\theta}) \right| d\theta \leq \frac{1}{r^n} \frac{r}{1-r}.$$

Thus,

$$|a_n| \leq \frac{1}{(1-r)r^{n-1}}.$$

To find the minimum value for $|a_n|$, we maximize $h(r) = (1-r)r^{n-1}$ in $[0, 1]$. From

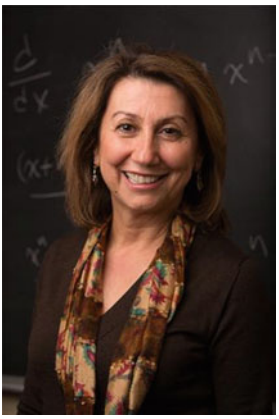
$$h'(r) = r^{n-2}((n-1)(1-r) - r),$$

we see the maximum attained when $r = 1 - \frac{1}{n}$. Therefore

$$|a_n| \leq \frac{1}{(1-r)r^{n-1}} = n\left(\frac{n}{n-1}\right)^{n-1} < e\, n. \qquad \square$$

In this section we proved several elementary results of univalent functions and normalized Schlicht classes using only knowledge from undergraduate analysis classes. In the decades that Bieberbach conjecture stood unsolved, mathematicians discovered many properties of univalent functions. In fact, most of the theory of univalent functions arose from partial results about $S$ and its subclasses. The significance of the Bieberbach conjecture does not lie only with its solution; but rather it belongs to the theory that was developed to solve it.

# About the Author



**Asuman Güven Aksoy** is Crown Professor of Mathematics at Clairmont McKenna College. Her research interests include functional analysis, metric geometry, and operator theory.

Professor Aksoy is coauthor of *A Problem Book in Real Analysis* (c) 2010 from the **Problem Books in Mathematics** series and *Nonstandard Methods in Fixed Point Theory* (c) 1990 in the **Universitext** series.

Additionally she is recipient of the MAA's Tensor Summa Grant 2013, the Fletcher Jones Grant for Summer Research 2009–2011, the MAA Award for Distinguished College or University teaching mathematics 2006, Huntoon Senior Teaching Award 2006, and the Roy P. Crocker Award for Merit, 2009, 2010.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

A. G. Aksoy, *Fundamentals of Real and Complex Analysis*, Springer Undergraduate Mathematics Series, https://doi.org/10.1007/978-3-031-54831-4

# Bibliography


1. S. Abbott. *Understanding analysis*. Undergraduate Texts in Mathematics. Springer-Verlag, New York, 2001.

2. J. Aczél. *Lectures on functional equations and their applications*. Mathematics in Science and Engineering, Vol. 19. Academic Press, New York-London, 1966. Translated by Scripta Technica, Inc. Supplemented by the author. Edited by Hansjorg Oser.

3. W. A. Adkins and S. H. Weintraub. *Algebra*, volume 136 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1992. An approach via module theory.

4. L. V. Ahlfors. An extension of Schwarz's lemma. *Trans. Amer. Math. Soc.*, 43(3):359–364, 1938.

5. L. V. Ahlfors. *Complex analysis: An introduction of the theory of analytic functions of one complex variable*. Second edition. McGraw-Hill Book Co., New York-Toronto-London, 1966.

6. A. G. Aksoy and M. A. Khamsi. *Nonstandard methods in fixed point theory*. Universitext. Springer-Verlag, New York, 1990. With an introduction by W. A. Kirk.

7. E. Artin. *The gamma function*. Translated by Michael Butler. Athena Series: Selected Topics in Mathematics. Holt, Rinehart and Winston, New York-Toronto-London, 1964.

8. S. Axler. *Measure, integration & real analysis*, volume 282 of *Graduate Texts in Mathematics*. Springer, Cham, [2020] ©2020.

9. J. A. Beachy and W. D. Blair. *Abstract Algebra*. Waveland Press, Inc., Long Grove, 2006.

10. R. Bernatz. *Fourier series and numerical methods for partial differential equations*. John Wiley & Sons, Inc., Hoboken, NJ, 2010.


© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

A. G. Aksoy, *Fundamentals of Real and Complex Analysis*, Springer Undergraduate Mathematics Series, https://doi.org/10.1007/978-3-031-54831-4

11. L. Bieberbach. Uber die koeffizienten derjenigen potenzreihen, welche eine schlichte abbildung des einheitskreises vermitteln. *Sitzungsberichte Akademie der Wissenschaften*, pages 940–955, 1916.

12. P. Bloomfield. *Fourier analysis of time series*. Wiley Series in Probability and Statistics: Applied Probability and Statistics. Wiley-Interscience [John Wiley & Sons], New York, 2000. An introduction, [2013] reprint of the second (2000) edition [MR1884963].

13. K. C. Border. *Fixed point theorems with applications to economics and game theory*. Cambridge University Press, Cambridge, 1989.

14. N. L. Carothers. *Real analysis*. Cambridge University Press, Cambridge, 2000.

15. E. W. Cheney. *Introduction to approximation theory*. AMS Chelsea Publishing, Providence, RI, 1998. Reprint of the second (1982) edition.

16. P. J. Cohen. *Set theory and the continuum hypothesis*. W. A. Benjamin, Inc., New York-Amsterdam, 1966.

17. C. W. Curtis. *Linear algebra*. Undergraduate Texts in Mathematics. Springer-Verlag, New York, fourth edition, 1993. An introductory approach.

18. K. R. Davidson and A. P. Donsing. *Real Analysis with Real Applications*. Prentice Hall, 2002.

19. L. de Branges. A proof of the Bieberbach conjecture. *Acta Math.*, 154(1–2):137–152, 1985.

20. P. Duren. *Invitation to classical analysis*, volume 17 of *Pure and Applied Undergraduate Texts*. American Mathematical Society, Providence, RI, 2012.

21. G. B. Folland. *Fourier analysis and its applications*. The Wadsworth & Brooks/Cole Mathematics Series. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 1992.

22. G. B. Folland. *Real analysis*. Pure and Applied Mathematics (New York). John Wiley & Sons, Inc., New York, second edition, 1999. Modern techniques and their applications, A Wiley-Interscience Publication.

23. K. Goebel. *Concise course on fixed point theorems*. Yokohama Publishers, Yokohama, 2002.

24. S. Gong. *The Bieberbach conjecture*, volume 12 of *AMS/IP Studies in Advanced Mathematics*. American Mathematical Society, Providence, RI; International Press, Cambridge, MA, 1999. Translated from the 1989 Chinese original and revised by the author, With a preface by Carl H. FitzGerald.

25. A. Hatcher. *Algebraic topology*. Cambridge University Press, Cambridge, 2002.

26. M. Hoffman and M. J. E. *Elementary Classical Analysis*. W. H. Freeman and Company, 1993.

27. B. B. Hubbard. *The world according to wavelets*. A K Peters, Ltd., Wellesley, MA, second edition, 1998. The story of a mathematical technique in the making.

28. T. W. Hungerford. *Algebra*, volume 73 of *Graduate Texts in Mathematics*. Springer-Verlag, New York-Berlin, 1980. Reprint of the 1974 original.

29. T. Jech. *Set theory*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2003. The third millennium edition, revised and expanded.

30. S. Katok. *p-adic analysis compared with real*, volume 37 of *Student Mathematical Library*. American Mathematical Society, Providence, RI; Mathematics Advanced Study Semesters, University Park, PA, 2007.

31. T. Kawata. *Fourier analysis in probability theory*. Academic Press, New York-London, 1972. Probability and Mathematical Statistics, No. 15.

32. M. A. Khamsi and W. A. Kirk. *An introduction to metric spaces and fixed point theory*. Pure and Applied Mathematics (New York). Wiley-Interscience, New York, 2001.

33. T. W. Körner. *Fourier analysis*. Cambridge University Press, Cambridge, 1988.

34. S. G. Krantz. *Geometric function theory*. Cornerstones. Birkhäuser Boston, Inc., Boston, MA, 2006. Explorations in complex analysis.

35. E. Kreyszig. *Introductory functional analysis with applications*. John Wiley & Sons, New York-London-Sydney, 1978.

36. M. Kuczma. *An introduction to the theory of functional equations and inequalities*, volume 489. Uniwersytet Ślaski, Katowice; Państwowe Wydawnictwo Naukowe (PWN), Warsaw, 1985. Cauchy's equation and Jensen's inequality, With a Polish summary.

37. S. Lang. *Linear algebra*. Undergraduate Texts in Mathematics. Springer-Verlag, New York, third edition, 1987.

38. G. G. Lorentz. *Bernstein polynomials*. Mathematical Expositions, no. 8. University of Toronto Press, Toronto, 1953.

39. J. Pawlikowski. The Hahn-Banach theorem implies the Banach-Tarski paradox. *Fund. Math.*, 138(1):21–22, 1991.

40. A. Pinkus. Weierstrass and approximation theory. *J. Approx. Theory*, 107(1):1–66, 2000.

41. H. A. Priestley. *Introduction to complex analysis*. Oxford University Press, Oxford, second edition, 2003.

42. W. Rudin. *Principles of mathematical analysis*. McGraw-Hill Book Co., New York-Auckland-Düsseldorf, third edition, 1976. International Series in Pure and Applied Mathematics.

43. P. J. Sally, Jr. *Tools of the trade*. American Mathematical Society, Providence, RI, 2008. Introduction to advanced mathematics.

44. Y. A. Shashkin. *Fixed points*, volume 2 of *Mathematical World*. American Mathematical Society, Providence, RI; Mathematical Association of America, Washington, DC, 1991. Translated from the Russian by Viktor Minachin [V. V. Minakhin].

45. C. G. Small. *Functional equations and how to solve them*. Problem Books in Mathematics. Springer, New York, 2007.

46. M. Spivak. *Calculus on manifolds. A modern approach to classical theorems of advanced calculus*. W. A. Benjamin, Inc., New York-Amsterdam, 1965.

47. G. Springer. *Introduction to Riemann surfaces*. Addison-Wesley Publishing Co., Inc., Reading, MA, 1957.

48. J. M. Steele. *The Cauchy-Schwarz master class*. MAA Problem Books Series. Mathematical Association of America, Washington, DC; Cambridge University Press, Cambridge, 2004. An introduction to the art of mathematical inequalities.

49. E. M. Stein and R. Shakarchi. *Fourier analysis*, volume 1 of *Princeton Lectures in Analysis*. Princeton University Press, Princeton, NJ, 2003. An introduction.

50. E. M. Stein and R. Shakarchi. *Real analysis*. Princeton Lectures in Analysis, III. Princeton University Press, Princeton, NJ, 2005. Measure theory, integration, and Hilbert spaces.

51. M. H. Stone. The generalized Weierstrass approximation theorem. *Math. Mag.*, 21:167–184, 237–254, 1948.

52. W. Wade. *An Introduction to Analysis*. Prentice Hall, 1999.

53. S. Wagon. *The Banach-Tarski paradox*, volume 24 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1985. With a foreword by Jan Mycielski.

54. D. G. Zill and P. D. Shanahan. *A first course in Complex Analysis with Applications (second edition)*. Jones and Bartlett Publishers, 2009.

55. P. Zorn. The Bieberbach conjecture. *Math. Mag.*, 59(3):131–148, 1986.

56. A. Zygmund. *Trigonometric series. Vol. I, II.* Cambridge Mathematical Library. Cambridge University Press, Cambridge, third edition, 2002. With a foreword by Robert A. Fefferman.

# Index

absolute convergence, 58
accumulation point of a set, 67
algebraic number, 22, 29
almost everywhere, 250
angle between curves, 315
annulus, 280
Archimedean property, 33
arithmetic-geometric mean inequality, 154
Arzela-Ascoli theorem, 144
axiom of choice, 18
axiom of completeness, 30

Baire's category theorem, 144
Banach space, 169
Banach-Tarski paradox, 272
Beppo-Levi theorem, 266
Bernoulli's inequality, 22
Bernstein polynomial, 206
Bessel's inequality, 230
Bieberbach conjecture, 375
Binomial coefficient, 207
Bolzano-Weierstrass theorem, 48, 74, 143
boundary of a set, 70
bounded convergence theorem, 259
bounded sequence, 48
bounded set, 30

Cantor-Lebesgue function, 252, 254
Cantor's diagonalization, 15
Cantor set, 77, 120
cardinality, 10
Cauchy condensation test, 57
Cauchy criterion, 50
Cauchy initial value problem, 183
Cauchy-Riemann equations, 290
Cauchy-Schwarz inequality, 152, 154
Cauchy sequence, 49, 134
Cauchy's formula for derivatives, 353
Cauchy's functional equation, 215
Cauchy's inequality, 354
Cauchy's integral formula, 350
Cauchy's theorem, 338
Cesàro sums, 233
characteristic function, 243
closed ball, 131
closure, 68
compact set, 70, 93, 143
comparison test, 59
complete metric space, 135
completeness, 30
complex exponential function, 298
complex fundamental theorem of calculus, 334
complex inner product, 149
complex logarithmic function, 307
complex numbers, 277

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2024

A. G. Aksoy, *Fundamentals of Real and Complex Analysis*, Springer Undergraduate Mathematics Series, https://doi.org/10.1007/978-3-031-54831-4

complex sequence, 53
conformal mapping, 315
connected sets, 75
continuous function, 87
continuum hypothesis, 15
contour, 338
contraction mapping, 175
contraction mapping theorem, 175
convergent sequence, 39, 134
convex functions, 221
convolution, 236
countable, 12
countably infinite, 12

d'Alembert's wave equation, 223
deformation, 339
de Moivre's formula, 285
De Morgan's Laws, 21, 145
dense, 33, 35, 121
differentiable function, 98
Dini's theorem, 197
Dirichlet formula, 232
Dirichlet function, 89, 112, 113, 241, 242
Dirichlet kernel, 231
Dirichlet problem, 320
Dirichlet test, 199
disk of convergence, 297
divergence test, 56
dominated convergence theorem, 268

Egorov's theorem, 252
entire functions, 298, 352
equicontinuous, 98, 144
equivalence class, 17
equivalence relation, 17
essential singularity, 364
Euler's formula, 279
exponential function, 304
extreme value theorem, 93

Fatou lemma, 264
Fejér kernel, 233
Fibonacci sequence, 38
field axioms, 25
finite intersection property, 74
fixed point, 174
Fourier coefficients, 226
Fourier series, 226
Fourier transform, 235
Fréchet metric, 162
fractional dimension, 81
Fredholm integral equation, 185
Frobenius norm, 166
Fubini's theorem, 236
function
    bijective, 7
    holomorphic, 289
    measurable, 247
    one-to-one, 21, 104, 105
    step, 124
    surjective, 6
functional equations, 213
fundamental theorem of algebra, 284, 352
fundamental theorem of calculus, 117, 119

gamma function, 213
Gauss' mean value theorem, 358
geometric series, 56
Gram-Schmidt procedure, 157
greatest lower bound, 31
Green's theorem, 337

Hadamard's formula, 297
Hamel basis, 217, 222
harmonic, 321
harmonic series, 57
Hausdorff maximality principle, 20
Heine-Borel theorem, 73
Hilbert space, 161
Hölder's Inequality, 129
holomorphic function, 290
homotopic, 339

identity map, 7
induction proof, 8
infinite series, 55
inner product, 150
inner product space, 147

integral
    Lebesgue, 255
    Riemann, 124
integral equations, 185
integration along paths, 331
interior of a set, 68
interior point, 65
intermediate value theorem, 96
intermediate value theorem for integrals, 116
inverse function theorem, 104, 108
isolated point of a set, 67
isolated singularity, 364
isometric, 137

jacobian, 292
Jensen's equation, 219
Jordan curve theorem, 338

kernel, 185
Koebe function, 376
Koebe $\frac{1}{4}$ - theorem, 380

Lagrange's identity, 289
Laplace's equation, 320
Laplacian, 237
Laurent series, 359
least upper bound, 30
Lebesgue criterion, 120
Lebesgue integral, 260
Lebesgue measurable, 246
Lebesgue's theorem, 122
Legendre polynomial, 159
lim inf, lim sup, 51
limit of a function, 84
linear approximation, 99
linear fractional transformations, 315
linear functional, 159
Liouville's theorem, 351
Lipschitz condition, 183
Littlewood's three principles, 251
local maximum, 101
local minimum, 101
Lusin's theorem, 252

matrix norms, 165
maximum modulus theorem, 356
mean-square approximation, 228
mean value theorem, 102, 110
measure, 65
measure zero, 120
metric
    discrete, 129
    Euclidean, 129
    Fréchet, 162
    taxicab, 131
    translation invariant, 162
    uniform, 131
metric space, 128
Minkowski's inequality, 155
Möbius transformation, 323
modes of convergence, 191
modular arithmetic, 26
Monotone bounded sequence, 44
monotone convergence theorem, 266
Morera's theorem, 353

nested sets theorem, 33, 75, 144
non-Archimedian triangle inequality, 138
norm, 45, 148

open and closed sets, 63
open ball, 131
open set, 63
operator norm, 168
order axioms, 24
orthogonal decompsition, 151
orthonormal basis, 156
orthonormal list, 156
Ostrowski's theorem, 141
outer measure, 244

p-adic absolute value, 140
p-adic numbers, 137
p-adic ordinal, 139
parallelogram equality, 164
partially ordered set, 19
Picard's theorem, 183
plane wave function, 237

pointwise convergence, 191
polar coordinates, 278
pole of order $m$, 364
power series, 296
power set, 3
projection, 18
Pythagorean theorem, 151

quotient set, 18
quotient space, 17

radius of convergence, 296
ratio test, 59
real numbers, 29
relation, 15
removable singularity, 364
residue, 366
residue theorem, 368
reverse triangle inequality, 35
Riemann integrable function, 112, 241
Riemann mapping theorem, 319
Riemann sphere, 287
Riemann sum, 111
Riemann surface, 311
Riesz representation theorem, 159
Rolle's theorem, 102
roots of unity, 285
root test, 60
Russell's paradox, 2

Schröder-Bernstein theorem, 11
Schwarz' Lemma, 357
second category, 147
sector, 281
separable, 203
sequence space, 164
sequentially compact, 143
set
    closed, 65, 131
    dense, 146
    diameter, 173
    empty, 21
    nowhere dense, 146
    open, 63, 131
simple function, 250
simply connected, 319
singularities, 339, 364
smooth functions, 101
square integrable, 228
subsequence, 47
support of a function, 258
symmetric difference of sets, 252
systems of algebraic linear equations, 179

tangent line, 110
Taylor's theorem, 355
Tchebychev inequality, 269
transcendental number, 29
triangle inequality, 35, 128, 137, 153
triplet representation of Mobius transformation, 325

ultrametric space, 139, 142
uncountable, 14
uniform Cauchy sequence, 199
uniform convergence, 192
    continuity, 193
    differentiation, 196
    integration, 195
    series, 197
uniformly continuous function, 95
uniformly differentiable, 110
univalent function, 375
upper lower sums, 112

Vitali's example of non-measurable set, 274
Volterra integral equation, 186

Weierstrass approximation theorem, 206, 229
Weierstrass function, 200
Weierstrass M-test, 198, 231
well ordering principle, 19, 20

Zahlen, 24
Zorn's lemma, 19